

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)



Europäisches Patentamt

Europ ean Patent Office

Office européen des brevets

(11) Publication number:

0 204 590

A2

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: 86400855.2

(51) Int. Cl. 4: C 12 N 15/00
A 01 H 1/00

(22) Date of filing: 21.04.86

(30) Priority: 22.04.85 US 725368

(43) Date of publication of application:
10.12.86 Bulletin 86/50

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: INSTITUT NATIONAL DE LA RECHERCHE
AGRONOMIQUE (INRA)
149, rue de Grenelle
F-75341 Paris Cedex 07(FR)

(72) Inventor: Slightom, Jerry L.
3305 Lorraine Avenue
Kalamazoo Michigan 49008(US)

(72) Inventor: Tepfer, David A.
9 Passage de la Main d'Or
F-75011 Paris(FR)

(74) Representative: Warcolin, Jacques et al.
Cabinet Régimbeau 26, avenue Kléber
F-75116 Paris(FR)

(54) T-DNA promoters of the Ri plasmid.

(57) The sequence of the T_L-DNA of Ri plasmids found in *Agrobacterium rhizogenes* strains HR1 and A4 is disclosed. Sixteen open reading frames bounded by eukaryotic promoters, ribosome binding sites, and polyadenylation sites were found, five of which were observed to be transcribed in a developmentally and phenotypically regulated manner. The use of promoters and polyadenylation sites from pRi T_L-DNA to control expression of heterologous foreign structural genes is taught, using as examples the structural genes for *Phaseolus vulgaris* storage protein (phaseolin), *P. vulgaris* lectin, a sweet protein (thaumatin), and *Bacillus thuringiensis* crystal protein. Vectors useful for manipulation of sequences of the structural genes and T-DNA are also provided.

EP 0 204 590 A2

TITLE MODIFIED
see front page Ri T-DNA PROMOTERS

FIELD

The present invention is in the fields of genetic engineering and plant husbandry, and especially provides means for promotion of transcription in plant.

BACKGROUND

Following are publications which disclose background information related to the present invention. These publications are discussed in greater depth in the Background sections indicated. Restriction maps of Ri plasmids are disclosed by G. A. Huffman et al. (1984) J. Bacteriol. 157:269-276; L. Jouanin (1984) Plasmid 12:91-102; and M. Pomponi et al. (1983) Plasmid 10:119-129 (see TIP Plasmid DNA). L. Herrera-Estrella et al. (1983) Nature 303:209-213, provides examples of use of the nos promoter to drive expression in plants of heterologous foreign structural genes. N. Murai et al. (1983) Science 222:476-482, reported the ocs promoter could drive expression of an intron-containing fusion gene having foreign coding sequences. (Manipulations of the TIP Plasmids). R. F. Barker et al. (1983) Plant Molec. Biol. 2:335-350, and R. F. Barker and J. D. Kemp, U.S. Patent application ser. no. 553,786 disclose the complete sequence of the T-DNA from the octopine-type plasmid pTi15955; homologous published sequences of other Ti plasmid genes are referenced therein. Barker and Kemp also taught use of various octopine T-DNA promoters to drive expression in plants of various structural genes (Genes on the TIP Plasmids).

Shuttle Vectors

Shuttle vectors, developed by G. B. Ruvkun and F. M. Ausubel (1981) Nature 289:85-88, which provide means for inserting foreign genetic material into large DNA molecules, include copies of recipient genome DNA sequences into which the foreign genetic material is inserted. Shuttle vectors can be introduced a recipient cell, by well known methods, inclu-

ding the tri-parental mating technique (Ruvkin and Ausubel, supra), direct transfer of a self-mobilizable vector in a bi-parental mating, direct uptake of exogenous DNA by Agrobacterium cells ("transformation"), spheroplast fusion of Agrobacterium with another bacterial cell, uptake of liposome-encapsulated DNA. After a shuttle vector is introduced into a recipient cell, possible events include a double cross-over with one recombinational event on either side of the marker (homogenotization). Phenotypically dominant traits may be introduced by single cross-over events (cointegration) (A. Caplan et al. (1983) Science 222:815-821; R. B. Horsch et al. (1984) Science 223:496-498); one must guard against deletion of the resulting tandem duplication. Shuttle vectors have proved useful in manipulation of Agrobacterium plasmids.

"Suicide Vectors" (e.g. R. Simon et al. (1983) Biotechnol. 1:784-791), are shuttle vectors having replicons not independently maintainable within the recipient cell. Use of suicide vectors to transfer DNA sequences into a Ti plasmid has been reported (e.g. E. Van Haute et al. (1983) EMBO J. 2:411-417; L. Comai et al. (1983) Plasmid 10:21-30; P. Zambryski et al. (1983) EMBO J. 2:2143-2150; P. Zambryski et al. (1984) in Genetic Engineering, Principles, and Methods, 6, eds: A. Hollaender and J. Setlow; P. Zahn et al. (1984) Mol. Gen. Genet. 194:188-194; and Caplan et al., supra; and C. H. Shaw et al. (1983) Gene 28:315-330.

Overview of Agrobacterium

Included within the gram-negative genus Agrobacterium are the species A. tumefaciens and A. rhizogenes, respectively the causal agents of crown gall disease and hairy root disease of gymnosperm and dicotyledonous angiosperm plants. In both diseases, the inappropriately growing plant tissue usually produces one or more amino acid derivatives known as opines which may be classified into families whose type members include octopine, nopaline, mannopine, and agropine.

Virulent strains of Agrobacterium harbor large plasmids known as Ti (tumor-inducing) plasmids (pTi) in A. tumefaciens and Ri (root-inducing) plasmids in A. rhizogenes (pRi), often classified by the opine which they caused to be synthesized. Ti and Ri plasmids both contain DNA sequences, referred to as T-DNA (transferred-DNA), which in tumors are found to be

integrated into the genome of the host plant. Several T-DNA genes are under control of T-DNA promoters which resembles the canonical eukaryotic promoter in structure. The Ti plasmid also carries genes outside the T-DNA region. The set of genes and DNA sequences responsible for transforming the plant cell are hereinafter collectively referred to as the transformation-inducing principle (TIP). The term TIP therefore includes, but is not limited to, both Ti and Ri plasmids.

5 General reviews of Agrobacterium-caused disease include those by D. J. Merlo (1982), *Adv. Plant Pathol.* 1:139-178; L. W. Ream and M. P. Gordon (1982), *Science* 218:854-859; M. W. Bevan and M.-D. Chilton (1982), *Ann. Rev. Genet.* 16:357-384; G. Kahl and J. Schell (1982) Molecular Biology of Plant Tumors; K. A. Barton and M.-D. Chilton (1983) *Meth. Enzymol.* 101:527-539; A. Depicker *et al.* (1983) in Genetic Engineering of Plants: an Agricultural Perspective, eds: T. Kosuge *et al.*, pp. 143-176; 10 A. Caplan *et al.* (1983) *Science* 222:815-821; T. C. Hall *et al.*, European Patent application 126,546; and A. N. Binns (1984) *Oxford Surveys Plant Mol. Cell Biol.* 1:130-160. A number of more specialized reviews can be found in A. Pühler, ed. (1983) Molecular Genetics of the Bacteria-Plant Interaction, including a treatment by D. Tepfer of A. rhizogenes-mediated transformation (pp. 248-258). R. A. Schilperoort (1984) in Efficiency in Plant Breeding (Proc. 10th Congr. Eur. Assoc. Res. Plant Breeding), eds: W. Lange *et al.*, pp. 251-285, discusses the Agrobacterium-based plant transformation in the context of the art of plant genetic engineering and plant improvement.

25

Infection of Plant Tissues

Plant cells can be transformed by Agrobacterium by several methods known to the art. For a review of recent work, see K. Syono (1984) *Oxford Surveys Plant Mol. Cell Biol.* 1:217-219. In the present invention, any 30 method will suffice as long as the gene is stably transmitted through mitosis and meiosis.

35

The infection of plant tissue by Agrobacterium is a simple technique well known to those skilled in the art. Typically after being wounded, a plant is inoculated with a suspension of tumor-inducing bacteria. Alternatively, tissue pieces are inoculated, e.g. leaf disks (R. B. Horsch

et al. (1985) Science 227:1229-1231) or inverted stem segments (K. A. Barton et al. (1983) Cell 32:1033-1043). After induction, the tumors can be placed in tissue culture on media lacking phytohormones usually included for culture of untransformed plant tissue. Traditional inoculation and culture techniques may be modified for use of disarmed T-DNA vectors incapable of inducing hormone independent growth (e.g. see P. Zambryski et al. (1984) in Genetic Engineering, Principles, and Methods, 6, eds.: A. Hollaender and J. Setlow).

Agrobacterium is also capable of infecting isolated cells, cells grown in culture, callus cells, and isolated protoplasts (e.g. R. B. Horsch and R. T. Fraley (1983) in Advances in Gene Technology: Molecular Genetics of Plants and Animals (Miami Winter Symposium 20), eds.: K. Downey et al., p. 576; R. T. Fraley et al. (1984) Plant Mol. Biol. 3:371-378; R. T. Fraley and R. B. Horsch (1983) in Genetic Engineering of Plants: an Agricultural Perspective, eds.: T. Kosuge et al., pp. 177-194; A. Muller et al. (1983) Biochem. Biophys. Res. Comm. 123:458-462). The transformation frequency of inoculated callus pieces can be increased by addition of an opine or opine precursors (L. M. Cello and W. L. Olsen, U.S. Patent 4,459,355).

Plant protoplasts can be transformed by the direct uptake of TIP DNA in the presence of a polycation, polyethelene glycol, or both (e.g. F. A. Krens et al. (1982) Nature 296:72-74), though integrated Ti plasmid may include non-T-DNA sequences.

An alternative method involves uptake of DNA surrounded by membranes. pTi-DNA may be introduced via liposomes or by fusion of plant and bacterial cells after removal of their respective cell walls (e.g. R. Hain et al. (1984) Plant Cell Rept. 3:60-64). Plant protoplasts can take up cell wall delimited Agrobacterium cells. T-DNA can be transmitted to tissue regenerated from fused protoplasts.

The host range of crown gall pathogenesis may be influenced by T-DNA-encoded functions such as onc genes (A. Hoekema et al. (1984) J. Bacteriol. 158:383-385; A. Hoekema et al. (1984) EMBO J. 3:3043-3047; W. C. Buchholz and M. F. Thomasshow (1984) 160:327-332). R. L. Ausich, European Patent Application 108,580, reports transfer of T-DNA from A. tumefaciens to green algal cells, and expression therein of octopine

synthase and Tn5 kanamycin resistance genes. G. M. S. Hooykaas-van Slogteren et al. (1984) Nature 311:763-764, and J.-P. Hernalsteens et al. (1984) EMBO J. 3:3039-3041, have demonstrated transformation of monocot cells by Agrobacterium without the customary tumorigenesis.

5 Regeneration of Plants

Differentiated plant tissues with normal morphology have been obtained from crown gall tumors. For example, L. Otten et al. (1981) Molec Gen. Genet. 183:209-213, used tms (shoot-inducing, root-suppressing) Ti plasmid mutants to create tumors which proliferated shoots that formed 10 self-fertile flowers. The resultant seeds germinated into plants which contained T-DNA and made opines. The tms phenotype can be partly overcome by washing of the rooting area and can be bypassed by grafting onto a normal stock (A. Wöstemeyer et al. (1984) Mol. Gen. Genet. 194:500-507). Similar experiments with a tmr (root-inducing, shoot-suppressing) mutant 15 showed that full-length T-DNA could be transmitted through meiosis to progeny and that in those progeny nopaline genes could be expressed, though at variable levels (K. A. Barton et al. (1983) Cell 32:1033-1043).

Genes involved in opine anabolism were capable of passing through 20 meiosis, though the plants were male sterile if the T-DNA was not disarmed. Seemingly unaltered T-DNA and functional foreign genes can be inherited in a dominant, closely linked, Mendelian fashion. Genetically, T-DNA genes are closely linked in regenerated plants (A. Wöstemeyer et al. (1984) Mol. Gen. Genet. 194:500-507; R. B. Horsch et al. (1984) Science 25 223:496-498; D. Tepfer (1984) Cell 37:959-967).

The epigenetic state of the plant cells initially transformed can affect regeneration potential (G. M. S. van Slogteren et al. (1983) Plant Mol. Biol. 2:321-333).

Roots resulting from transformation from A. rhizogenes have proven 30 relatively easy to regenerate directly into plantlets (M.-D. Chilton et al. (1982) Nature 295:432-434; D. Tepfer (1984) Cell 37:959-967; Tepfer (1983) in Puhler, supra), and are easily cloned. Regenerability from transformed roots may be dependent on T-DNA copy-number (C. David et al. (1984) Biotechnol. 2:73-76). Hairy root regenerants have a rhizogenic

- 6 -

potential and isozyme pattern not found in untransformed plants (P. Costantino et al. (1984) *J. Mol. Appl. Genet.* 2:465-470). The phenotype of these plants is generally altered, although not necessarily deleteriously.

5 Genes on the Ti Plasmids

The complete sequence of the T-DNA of an octopine-type plasmid found in ATCC 15955, pTi15955, has been reported (R. F. Barker et al. (1983) *Plant Molec. Biol.* 2:335-350), as has that of the T_L region of pTiAch5 (J. Gielen et al. (1984) *EMBO J.* 3:835-846). Published T-DNA genes do not contain introns and do have sequences that resemble canonical eukaryotic promoter elements and polyadenylation sites.

10 Ti plasmids having mutations in the genes tms, tmr, tml, and ocs respectively incite tumorous calli of Nicotiana tabacum which generate shoots, proliferate roots, are larger than normal, and do not synthesize octopine; all but ocs are onc (oncogenicity) genes. In other hosts, mutants of these genes can induce different phenotypes (see M. W. Bevan and M.-D. Chilton (1982) *Ann. Rev. Genet.* 16:357-384). Mutations in T-DNA genes do not seem to affect the insertion of T-DNA into the plant genome (J. Leemans et al. (1982) *EMBO J.* 1:147-152; L. W. Ream et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:1660-1664).

15 Octopine Ti plasmids carry an ocs gene which encodes octopine synthase (lysopine dehydrogenase). All upstream signals necessary for expression of the ocs gene are found within 295 bp of the ocs transcriptional start site (C. Koncz et al. (1983) *EMBO J.* 2:1597-1603). P. Dhaese et al. (1983) *EMBO J.* 2:419-426, reported the utilization of various polyadenylation sites by "transcript 7" (ORF3 of Barker et al., supra) and ocs. The presence of the enzyme octopine synthase within a tissue can protect that tissue from the toxic effect of various amino acid analogs (G. A. Dahl and J. Tempé (1983) *Theor. Appl. Genet.* 66:233-239; M. G. Koziel et al. (1984) *J. Mol. Appl. Genet.* 2:549-562).

20 Nopaline Ti plasmids encode the nopaline synthase gene (nos) (sequenced by A. Depicker et al. (1982) *J. Mol. Appl. Genet.* 1:561-573). The "CAAT" box, but not upstream sequences therefrom, is required for wild-type levels of nos expression; a partial or complete "TATA" box

25

supports very low level nos activity (C. H. Shaw et al. (1984) Nucl. Acids Res. 12:7831-7846). Genes equivalent to tms and tmr have been identified on a nopaline-type plasmid and a number of transcripts have been mapped (L. Willmitzer et al. (1983) Cell 32:1045-1056).

5 Transcription from hairy root T-DNA has also been detected (L. Willmitzer et al. (1982) Mol. Gen. Genet. 186:16-22). Ri plasmids and tms- Ti plasmids can complement each other when inoculated onto plants, resulting in calli capable of hormone-independent growth (G. M. S. van Slogteren (1983) Ph.D. thesis, Rijksuniversiteit te Leiden, 10 Netherlands).

TIP plasmid genes outside of the T-DNA region include the vir genes, which when mutated result in an avirulent Ti plasmid. Several vir genes have been accurately mapped and have been found to be located in regions conserved among various Ti plasmids (V. N. Iyer et al. (1982) Mol. Gen. Genet. 188:418-424). The vir genes function in trans, being capable of causing the transformation of plant cells with T-DNA of a different plasmid type and physically located on another plasmid (e.g. A. J. de Framond et al. (1983) Biotechnol. 1:262-269; A. Hoekema et al. (1983) Nature 303:179-180; J. Hille et al. (1984) J. Bacteriol. 158:754-756; 15 A. Hoekema et al. (1984) J. Bacteriol. 158:383-385); such arrangements are known as binary systems. Chilton et al. (18 January 1983) 15th Miami 20 Winter Symp., described a "micro-Ti" plasmid made by resectioning the "mini-Ti" of de Framond et al., supra (see European Patent application 126,546 for a description). G. A. Dahl et al., U.S. Patent application 25 ser. no. 532,280, and A. Hoekema (1985) Ph.D. Thesis, Rijksuniversiteit te Leiden, The Netherlands, disclose micro-Ti plasmids carrying ocs genes constructed from pTi15955. M. Bevan (1984) Nucl. Acids Res. 12:8711-8721, discloses a kanamycin-resistant micro-Ti. T-DNA need not be on a plasmid 30 to transform a plant cell; chromosomally located T-DNA is functional (A. Hoekema et al. (1984) EMBO J. 3:2485-2490). Ti plasmid-determined characteristics have been reviewed by Merlo, supra (see especially Table II therein), and Ream and Gordon, supra.

TIP Plasmid DNA

Ri plasmids have been shown to have extensive homology among themselves (P. Costantino *et al.* (1981) *Plasmid* 5:170-182), and to both octopine (F. F. White and E. W. Nester (1980) *J. Bacteriol.* 144:710-720) and nopaline (G. Risuleo *et al.* (1982) *Plasmid* 7:45-51) Ti plasmids, primarily in regions encoding vir genes, replication functions, and opine metabolism functions (L. Jouanin (1984) *Plasmid* 12:91-102; K. Lahners *et al.* (1984) *Plasmid* 11:130-140; E. E. Hood *et al.* (1984) *Biotechnol.* 2:702-709; F. Leach (1983) Ph.D. Thesis, Universite de Paris-Sud, Centre d'Orsay, France); none of the homologies are in pRi T_L-DNA. pRi T-DNA contains extensive though weak homologies to T-DNA from both types of Ti plasmid (L. Willmitzer *et al.* (1982) *Mol. Gen. Genet.* 186:16-22). DNA from several plant species contains sequences, referred to as cT-DNA (cellular T-DNA), having homology with the Ri plasmid (F. F. White *et al.* (1983) *Nature* 301:348-350, L. Spano *et al.* (1982) *Plant Molec. Biol.* 1:291-300; D. Tepfer (1982) in 2e Colloque sur les Recherches Fruitieres Bordeaux, pp. 47-59). G. A. Huffman *et al.* (1984) *J. Bacteriol.* 157:269-276 and Jouanin, *supra*, and Leach, *supra*, have shown that, in the region of cross-hybridization, the Ri plasmid pRiA4_b is more closely related to a pTiA6 (octopine-type) than pTiT37 (nopaline-type) and that this Ri plasmid appears to carry sequence homologous to tms but not tmr. Their results also suggested that Ri T-DNA may be discontinuous, analogous to the case with octopine T-DNA (see below). The restriction maps of pRiA4_b, pRi1855, and pRiHRI were respectively disclosed by Huffman *et al.*, *supra*, M. Pomponi *et al.* (1983) *Plasmid* 10:119-129, and L. Jouanin *supra*. Ri plasmids are often characterizable as being agropine-type or mannopine-type (A. Petit *et al.* (1983) *Mol. Gen. Genet.* 190:204-214).

A portion of the Ti or Ri plasmid is found in the DNA of tumorous plant cells. T-DNA may be integrated (i.e. inserted) into host DNA at multiple sites in the nucleus. Flanking plant DNA may be either repeated or low copy number sequences. Integrated T-DNA can be found in either direct or inverted tandem arrays and can be separated by spacers. Much non-T-DNA Ti plasmid DNA appears to be transferred into the plant cell prior to T-DNA integration (H. Joos *et al.* (1983) *EMBO J.* 2:2151-2160). T-DNA has direct repeats of about 25 base pairs associated with the

borders, i.e. with the T-DNA/plant DNA junctions, which may be involved in either transfer from Agrobacterium or integration into the host genome.

R_f plasmids integrate two separate T-DNAs, T_L-DNA and T_R-DNA, left and right T-DNAs, respectively. T_L (about 15-20 kbp) and T_R (about 6-10 kbp) are separated by about 15-20 kbp (Huffman et al., supra, Jouanin, supra). The region of agropine-type pR1 T_L and T_R integrated can vary between individual plants or species inoculated (F. F. White et al. (1983) *Nature* 301:348-350; D. A. Tepfer (1984) *Cell* 37:959-967). Though T-DNA is occasionally deleted after integration in the plant genome, it is generally stable. Tumors containing a mixture of cells which differ in T-DNA organization or copy number are the result of multiple transformation events.

The exact location relative to the border repeats of T-DNA/flanking plant DNA junctions varies and need not be within a border repeat. Virulence is not always eliminated after deletion of one of either of the usual nopaline T-DNA border sequences (compare H. Joos et al. (1983) *Cell* 32:1057-1067 with K. Wang et al. (1984) *Cell* 38:455-462 and C. H. Shaw et al. (1984) *Nucl. Acids Res.* 12:6031-6041, concerning the right border). The orientation of the right nopaline border can be reversed without total loss of functionality, and a single border sequence is capable of transforming closely-linked sequences (M. De Block et al. (1984) *EMBO J.* 3:1681-1689). A synthetic 25 bp nopaline right border repeat is functional (Wang et al., supra). Circular intermediates associated with T-DNA transfer appear to be spliced precisely within the 25 bp direct repeats (Z. Koukolikova-Nicola et al. (1985) *Nature* 313:191-196).

Manipulations of the TIP Plasmids

Altered DNA sequences, including deletions, may be inserted into TIP plasmids (see Shuttle Vectors). Some pTi derivatives can be transferred to E. coli and mutagenized therein (J. Hille et al. (1983) *J. Bacteriol.* 154:693-701). P. Zambryski et al. (1983) *EMBO J.* 2:2143-2150, report use of a vector, deleted for most T-DNA genes to transform tobacco and regenerate morphologically normal plants.

- 10 -

The nopaline synthase promoter can drive expression of drug resistance structural genes useful for selection of transformed plant cells. M. W. Bevan et al. (1983) *Nature* 304:184-187; R. T. Fraley et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:4803-4807; and L. Herrera-Estrella et al. (1983) *EMBO J.* 2:987-995, have inserted the bacterial kanamycin resistance structural gene (neomycin phosphotransferase II, NPT2), or kan, from Tn5 downstream from (i.e. behind or under control of) the nopaline synthase promoter. The constructions were used to transform plant cells which in culture were resistant to kanamycin and its analogs such as neomycin and G418. Promoters for octopine T_L genes ORF24 and ORF25 can also drive kan structural gene expression (J. Velten et al. (1984) *EMBO J.* 3:2723-2730). Herrera-Estrella et al., supra, reported a similar construction, in which a methotrexate resistance gene (dihydrofolate reductase, DHFR) from Tn7 was placed behind the nos promoter; transformed plant cells were resistant to methotrexate. Furthermore, L. Herrera-Estrella et al. (1983) *Nature* 303:209-213, have obtained expression in plant cells of enzymatic activity of octopine synthase and chloramphenicol acetyltransferase by placing their structural genes under control of nos promoters. G. Helmer et al. (1984) *Biotechnol.* 2:520-527, have created a fusion gene useful as a screenable marker having the promoter and 5'-end of the nos structural gene fused to E. coli β -galactosidase (lacZ) sequences.

N. Murai et al. (1983) *Science* 222:476-482, reported fusion of the promoter and the 5'-end of the octopine synthase structural gene to a phaseolin structural gene. The encoded fusion protein was produced under control of the T-DNA promoter. Phaseolin-derived introns underwent proper post-transcriptional processing.

SUMMARY OF THE INVENTION:

One object of this invention is to provide means for promoting the expression of structural genes within plant cells wherein said genes are foreign to said cells. In pursuance of this goal, other objects are to provide pRI T-DNA promoters and transcript terminators, and especially pRI T_L-DNA-derived promoters and pRI T_L-DNA-derived polyadenylation sites, which are DNA sequences capable of controlling structural gene transcrip-

tion and translation within plant cells, and to provide developmental and phenotypic regulation of said foreign structural genes. Another object is to provide specialized plant tissues and plants having within them proteins encoded by foreign structural genes and, in cases where the protein is an enzyme, having or lacking metabolites or chemicals which respectively are not or are otherwise found in the cells in which the genes is inserted. Other objects and advantages will become evident from the following description.

The invention disclosed herein provides a plant comprising a genetically modified plant cell having a foreign structural gene introduced and expressed therein under control of pRi T_L -DNA-derived plant expressible transcription controlling sequences (TxCS). Further, the invention provides plant tissue comprising a plant cell whose genome includes T-DNA comprising a foreign structural gene inserted in such orientation and spacing with respect to pRi T_L -DNA-derived plant-expressible TxCS as to be expressible in the plant cell under control of those sequences. Also provided are novel strains of bacteria containing and replicating T-DNA, the T-DNA being modified to contain an inserted foreign structural gene in such orientation and spacing with respect to a T-DNA-derived, plant-expressible TxCS as to be expressible in a plant cell under control of said TxCS. Additionally, the invention provides novel vectors having the ability to replicate in E. coli and comprising T-DNA, and further comprising a foreign structural gene inserted within T-DNA contained within the vector, in such manner as to be expressible in a plant cell under control of a pRi T_L -DNA TxCS. Furthermore, strains of bacteria harboring said vectors are disclosed.

Much is known about the location, size, and function of many transcripts activated when A. tumefaciens T-DNA regions are transferred into the genome of plants (see Background). Most pTi T-DNA T_L -DNA open reading frames (ORFs) correlate with known gene products. However, until the disclosure of the present invention, the art knew little about the number, size, and function of genes activated when the T_L -DNA regions from A. rhizogenes plasmids, such as pRiA4, are transferred into a plant genome. Agropine synthase, tms-1 and tms-2 genes have been identified by homology with pTi T-DNA in Ri plasmids, but these loci are located in

pRI T_R-DNA (G. A. Huffman *et al.* (1984) *J. Bacteriol.* 157:269-276; L. Jouanin (1984) *Plasmid* 12:91-102). The experimental work presented herein is believed to be the first disclosure of a pRI T_L-DNA sequence or of any sequence homologous thereto. The availability of this sequence 5 will enable and otherwise facilitate work in the art of plant transformation to express foreign structural genes and to engage in other manipulations of pRI T_L-DNA and pRI T_L-DNA-derived sequences. Without the newly disclosed pRI T_L-DNA sequence, those of ordinary skill in the art would be unable to use promoters and polyadenylation sites contained therein to 10 promote transcription and translation in plant cells of foreign structural genes. The disclosed sequence reveals the existence of previously unknown T-DNA ORFs and associated transcription controlling sequences, and makes possible construction of recombinant DNA molecules using promoters and 15 polyadenylation sites from pRI T_L-DNA genes whose sequences were hitherto unknown and unavailable to the public. The work presented herein is also believed to be the first disclosure of developmental and phenotypic regulation 20 of T-DNA genes. Results newly disclosed herein will allow those of ordinary skill in the art to use T-DNA transcription controlling sequences which are so regulated to express heterologous foreign structural genes in transformed plants. T-DNA genes known to the art before the present disclosure are not known to be so regulated. Furthermore, knowledge of pRI T_L-DNA sequence enables one to bring to utility promoters and polyadenylation sites that are presently unrecognized; in the future, should a 25 new pRI T_L-DNA transcript be discovered and mapped, the sequence disclosed herein will permit associated TxCSs to be combined with heterologous foreign structural genes.

The present invention comprises foreign structural genes under control of pRI T_L-DNA promoters expressible in plant cells, the promoter/gene combination being inserted into a plant cell by any means known to the 30 art. More specifically, in its preferred embodiment the invention disclosed herein comprises expression in plant cells of foreign structural genes under control of certain pRI T_L-DNA-derived plant expressible TxCSs, after introduction via T-DNA, that is to say, by inserting the foreign structural gene into T-DNA under control of a pRI T_L-DNA promoter and/or 35 ahead of a pRI T_L-DNA polyadenylation site and introducing the T-DNA containing the TxCS/structural gene combination into a plant cell using known

means. Once plant cells transformed to contain a foreign structural gene expressible under control of a pRI T_L-DNA TxCS are obtained, plant tissues and whole plants can be regenerated therefrom using methods and techniques well known in the art. The regenerated plants are then reproduced by conventional means and the introduced genes can be transferred to other strains and cultivars by conventional plant breeding techniques. The invention in principle applies to any introduction of a foreign structural gene combined with a pRI T_L-DNA promoter or polyadenylation site into any plant species into which foreign DNA (in the preferred embodiment pTi T-DNA) can be introduced and maintained by any means. In other words, the invention provides a means for expressing a structural gene in a plant cell and is not restricted to any particular means for introducing foreign DNA into a plant cell and maintaining the DNA therein. Such means include, but are not limited to, T-DNA-based vectors (including pTi-based vectors), viral vectors, minichromosomes, non-T-DNA integrating vectors, and the like.

The invention is useful for genetically modifying plant cells, plant tissues, and whole plants by inserting useful structural genes from other species, organisms, or strains that change phenotypes of plants or plant cells when expressed therein. Such useful structural genes include, but are not limited to, genes conveying phenotypes such as improved tolerance to extremes of heat or cold; improved tolerance to drought or osmotic stress; improved resistance or tolerance to insect (e.g. insecticidal toxins), arachnid, nematode, or epiphyte pests and fungal, bacterial, or viral diseases, or the like; the production of enzymes or secondary metabolites not normally found in said tissues or plants; improved nutritional (e.g. storage proteins or lectins), flavor (e.g. sweet proteins), or processing properties when used for fiber or human or animal food; changed morphological traits or developmental patterns (e.g. leaf hairs which protect the plant from insects, aesthetically pleasing coloring or form, changed plant growth habits, dwarf plants, reduced time needed for the plants to reach maturity, expression of a gene in a tissue or at a time that gene is not usually expressed, and the like); male sterility; improved photosynthetic efficiency (including lowered photorespiration); improved nitrogen fixation; improved uptake of nutrients; improved tolerance to herbicides; increased crop yield; improved competition with

other plants; and improved germplasm identification by the presence of one or more characteristic nucleic acid sequences, proteins, or gene products, or phenotypes however identified (to distinguish a genetically modified plant of the present invention from plants which are not so modified, to facilitate transfer of a linked artificially introduced phenotype by other (e.g. sexual) means to other genotypes or to facilitate identification of plants protected by patents or by plant variety protection certificates); selectable markers (i.e. genes conveying resistance in cell or tissue culture to selective agents); screenable markers; and the like.

The invention is exemplified by introduction and expression of a structural gene for phaseolin, the major seed storage protein of the bean Phaseolus vulgaris L., into plant cells. The introduction and expression of the structural gene for phaseolin, for example, can be used to enhance the protein content and nutritional value of forage or other crops. The invention is also exemplified by the introduction and expression of a lectin structural gene, in this case also obtained from P. vulgaris, into plant cells. The introduction and expression of a novel lectin may be used to change the nutritional or symbiotic properties of a plant tissue. The invention is exemplified in yet other embodiments by the introduction and expression of DNA sequences encoding thaumatin, and its precursors prothaumatin, prethaumatin, and preprothaumatin. Mature thaumatin is a heat-labile, sweet-tasting protein found naturally in katemfe (Thaumatooccus daniellii) which can be used to enhance the flavor of vegetables which are eaten uncooked without significantly increasing the caloric content of the vegetables. The invention is further exemplified by introduction and expression of a structural gene for a crystal protein from B. thuringiensis var. kurstaki HD-73 into plant cells. The introduction and expression of the structural gene for an insecticidal protein can be used to protect a crop from infestation with insect larvae of species which include, but are not limited to, hornworm (Manduca sp.), pink bollworm (Proctonophora gossypiella), European corn borer (Ostrinia nubilalis), tobacco budworm (Heliothis virescens), and cabbage looper (Trichoplusia ni). Applications of insecticidal protein prepared from sporulating B. thuringiensis does not control insects such as the pink bollworm in the field because of their particular life cycles and feeding habits. A plant containing in its tissues insecticidal protein will con-

5 trol this recalcitrant type of insect, thus providing advantage over prior insecticidal uses of B. thuringiensis. By incorporation of the insecticidal protein into the tissues of a plant, the present invention additionally provides advantage over such prior uses by eliminating instances of
10 nonuniform application and the costs of buying and applying insecticidal preparations to a field. Also, the present invention eliminates the need for careful timing of application of such preparations since small larvae are most sensitive to insecticidal protein and the protein is always present, minimizing crop damage that would otherwise result from preapplication larval foraging. Other uses of the invention, exploiting the properties of other structural genes introduced into various plant species, will be readily apparent to those skilled in the art.

15 DESCRIPTION OF THE DRAWINGS

Figure 1. presents maps of the T_L -DNA of agropine Ri plasmid pRiHRI and strategy used for sequencing. The top line represents the T_L -DNA region from pRiHRI and the filled boxes indicate locations of ORFs 1 to 20 18. The left and right T_L -DNA borders are those identified from analysis of T_L -DNA integrated into Convolvulus arvensis clone 7 tissue. ORF polarities are indicated by the position of enclosed boxes on the continuous line; above indicates transcription from left to right and below indicates transcription right to left, i.e. having an mRNA sequence complementary to that disclosed in Fig. 2. EcoRI and BamHI restriction maps are below the ORF map. The complete nucleotide sequence of the T_L -DNA was determined from five subclones mapped below the restriction maps:

25 EcoRI 3a, BamHI 8a; Number 16, pLJ0 ("cosmid 40"); and EcoRI 3b (see Example 2.2). Comparison of restriction enzyme site patterns (L. Jouanin 30 (1984) Plasmid 12:91-102) and overlapping nucleotide sequenced region (Number 16 and cosmid 40) indicate that pRiHRI and pRiA4 T_L -DNAs are essentially identical. Cleavage sites and direction of sequence analysis are shown below each subclone, and horizontal arrows indicate direction and distance of sequencing runs. Enzymes are abbreviated as follows:

35 A, AvaI; Ac, AccI; B, BamHI; Bg, BglII; C, ClaI; D, DraI; E, EcoRI; H, HindIII; K, KpnI; MsI, MstI; MsII, MstII; Na, NarI; Nc, NcoI; Ps, PstI; Pv, PvuII; Sa, SalI; St, StuI; Xb, XbaI; Xh, XhoI; Xm, XmnI; and Xo, XorII.

- 16 -

Figure 2. presents nucleotide sequence of T_L -DNA region from A. rhizogenes agropine-type plasmid pRiHRI. The sequence starts 520 base pairs (bp) to the left of the left T_L -DNA/plant junction sequence identified in C. arvensis clone 7 and extends 1135 bp to the right of the 5 clone 7 right T_L -DNA/plant junction, a total of 21,126 bp.

Figure 3. is a schematic diagram, not drawn to scale, of the DNA manipulation strategy utilized in the Examples. Sites susceptible to the 10 action of a restriction enzyme are indicated by that enzyme's name or place of listing in a Table. For example, "T4c2" refers to an enzyme listed in Table 4, column 2. A site that is no longer susceptible to the enzyme is indicated by the presence of parenthesis around the name of the 15 enzyme. The extent and polarity of an ORF is indicated by an arrow. Names of plasmids, again sometimes designated by place of listing in a Table (e.g. "T5c1" refers to a vector listed in Table 5, column 1), are within the circular representations of the plasmids. Names of vectors, again sometimes designated by a listing in a Table, are within the 20 circular representations of the plasmids. "Ex" refers to the Example which describes a particular manipulation.

20

DETAILED DESCRIPTION OF THE INVENTION

The following terms are defined in order to remove ambiguities to the intent or scope of their usage in the Specification and Claims.

25 TxCS: Transcription controlling sequences refers to a promoter/trans-
script terminator combination flanking a particular structural gene or
open reading frame (ORF). The promoter and transcript terminator DNA
sequences flanking a particular inserted foreign structural gene need not
be derived from the same source genes (e.g. pairing two different
30 pRi T_L -DNA) genes or the same taxonomic source (e.g. pairing sequences
from pRi T_L -DNA with sequences from non-pRi- T_L -DNA sources such as other
types of T-DNA, plants, animals, fungi, yeasts, and eukaryotic viruses).
Therefore the term TxCS refers to either combination of a claimed promoter
with an unclaimed transcript terminator, or combination of a unclaimed
35 promoter with a claimed polyadenylation site, or combination of a promoter
and a polyadenylation site which are both claimed. Examples of non-

pRI-T_L-DNA plant-expressible promoters which can be used in conjunction with a pRI T_L-DNA polyadenylation site include, but are not limited to, those from genes for nos, ocs, phaseolin, RuBP-Case small subunit and the 195 and 35S transcripts of cauliflower mosaic virus (CaMV).

5 Promoter: Refers to sequences at the 5'-end of a structural gene involved in initiation of translation or transcription. Expression under control of a pRI T-DNA promoter may take the form of direct expression in which the structural gene normally controlled by the promoter is removed in part or in whole and replaced by the inserted foreign structural gene, a start codon being provided either as a remnant of the pRI T-DNA structural gene or as part of the inserted structural gene, or by fusion protein expression in which part or all of the structural gene is inserted in correct reading frame phase within the existing pRI T-DNA structural gene. In the latter case, the expression product is referred to as a fusion protein. The promoter segment may itself be a composite of segments derived from a plurality of sources, naturally occurring or synthetic. Eukaryotic promoters are commonly recognized by the presence of DNA sequences homologous to the canonical form 5'...TATAA...3' about 10-30 bp 5' to the location of the 5'-end of the mRNA (cap site). About 30 bp 5' to the TATAA another promoter sequence is often found which is recognized by the presence of DNA sequences homologous to the canonical form 5'...CCAAT...3'. Translational initiation often begins at the first 5'...AUG...3' 3'-from the cap site (see Example 1.5).

10 20 25

30 35 Transcript terminator: Refers to any nucleic acid sequence capable of determining the 3'-end of a eukaryotic messenger RNA (mRNA). The transcript terminator DNA segment may itself be a composite of segments derived from a plurality of sources, naturally occurring or synthetic, and may be from a genomic DNA or an RNA-derived cDNA. Some eukaryotic RNAs, e.g. histone mRNA (P. A. Krieg and D. A. Melton (1984) *Nature* 308:203-206), ribosomal RNA, and transfer RNA, are not 3'-terminated by polyadenylic acid or by polyadenylation sites; it is intended that the term transcript terminator include, but not be limited to, both nucleic acid sequences determining the 3'-ends of such transcripts and polyadenylation site sequences (see below).

5 Polyadenylation site: Refers to any nucleic acid sequence capable of determining the 3'-end of a eukaryotic polyadenylated mRNA. After transcriptional termination polyadenylic acid "tails" are added to the 3'-end of most mRNA precursors. The polyadenylation site DNA segment may itself be a composite of segments derived from a plurality of sources, naturally occurring or synthetic, and may be from a genomic DNA or an mRNA-derived cDNA. Polyadenylation sites are commonly recognized by the presence of homology to the canonical form 5'...AATAAA...3', although variation of distance, partial "read-thru", and multiple tandem canonical sequences are not uncommon. It should be recognized that a canonical "polyadenylation site" may in fact not actually cause polyadenylation per se (N. Proudfoot (1984) Nature 307:412-413) and that sequences 3' to the "AATAAA" and the 3'-end of the transcript may be needed (A. Gil and N. J. Proudfoot (1984) Nature 312:473-474).

10

15 Foreign structural gene: As used herein includes that portion of a gene comprising a DNA segment coding for a foreign RNA, protein, polypeptide or portion thereof, possibly including a translational start codon, but lacking at least one other functional element of a TxCS that regulates initiation or termination of transcription and initiation of translation, 20 commonly referred to as the promoter region and transcript terminator. As used herein, the term foreign structural gene does not include pRI T_L -DNA structural genes unless the structural gene and pRI T_L -DNA transcription controlling sequences combined with the structural gene are derived from different pRI T_L -DNA genes; i.e. unless the structural gene and either a pRI promoter or a pRI polyadenylation site combined with the structural 25 gene are heterologous. (Note that such foreign functional elements may be present after combination of the foreign structural gene with a pRI T_L -DNA TxCS, though, in embodiments of the present invention, such elements may not be functional in plant cells). A foreign structural gene may encode a protein not normally found in the plant cell in which the gene is introduced. Additionally, the term refers to copies of a structural gene naturally found within the cell but artificially introduced. A foreign structural gene may be derived in whole or in part from sources including but not limited to eukaryotic DNA, prokaryotic DNA, episomal DNA, plasmid 30 DNA, plastid DNA, genomic DNA, cDNA, viral DNA, viral cDNA, or chemically synthesized DNA. It is further contemplated that a foreign structural 35

gene may contain one or more modifications in either the coding segments or untranslated regions which could affect the biological activity or chemical structure of the expression product, the rate of expression or the manner of expression control. Such modifications include, but are not limited to, mutations, insertions, deletions, and substitutions of one or more nucleotides, and "silent" modifications that do not alter the chemical structure of the expression product but which affect intercellular localization, transport, excretion or stability of the expression product. The structural gene may constitute an uninterrupted coding sequence or it may include one or more introns, bounded by the appropriate plant functional splice junctions, which may be obtained from synthetic or a naturally occurring source. The structural gene may be a composite of segments derived from a plurality of sources, naturally occurring or synthetic, coding for a composite protein, the composite protein being foreign to the cell into which the gene is introduced and expressed or being derived in part from a foreign protein. The foreign structural gene may be a fusion protein, and in particular, may be fused to all or part of a structural gene derived from the same ORF as was the TxCS.

Plant tissue: Includes differentiated and undifferentiated tissues of plants including, but not limited to roots, shoots, pollen, seeds, tumor tissue, such as crown galls, and various forms of aggregations of plant cells in culture, such as embryos and calluses. The plant tissue may be in planta or in organ, tissue, or cell culture.

Plant cell: As used herein includes plant cells in planta and plant cells and protoplasts in culture.

Production of a genetically modified plant, plant seed, plant tissue, or plant cell expressing a foreign structural gene under control of a pRI T-DNA TxCS, and especially a pRI T_L-DNA-derived TxCS, combines the specific teachings of the present disclosure with a variety of techniques and expedients known in the art. In most instances, alternative expedients exist for each stage of the overall process. The choice of expedients depends on variables such as the choice of the basic vector system for the introduction and stable maintenance of the pRI T_L-DNA TxCS/structural gene combination, the plant species to be modified and the desired regeneration strategy, and the particular foreign structural gene

to be used, all of which present alternative process steps which those of ordinary skill are able to select and use to achieve a desired result. For instance, although the starting point for obtaining pRi T_L-DNA TxCSs is exemplified in the present application by pRi T_L-DNA isolated from pRiA4 and pRiHRI, DNA sequences of other homologous agropine-type Ri Ti plasmids might be substituted as long as appropriate modifications are made to the TxCS isolation and manipulation procedures. Additionally, T-DNA genes from other types of pRi T_L-DNA homologous to the agropine-type pRi T_L-DNA genes having TxCSs disclosed herein may be substituted, again with appropriate modifications of procedural details. Homologous genes may be identified by those of ordinary skill in the art by the ability of their nucleic acids to cross-hybridize under conditions of stringency appropriate to detect 70% homology; such conditions are well understood in the art. It will be understood that there may be minor sequence variations within gene sequences utilized or disclosed in the present application. These variations may be determined by standard techniques to enable those of ordinary skill in the art to manipulate and bring into utility the T-DNA promoters and transcript terminators of such homologous genes. (Homologs of foreign structural genes may be identified, isolated, sequenced, and manipulated as is in a similar manner as homologs of the pRi genes of the present invention.) As novel means are developed for the stable insertion of foreign genes in plant cells, those of ordinary skill in the art will be able to select among those alternate process steps to achieve a desired result. The fundamental aspects of the invention are the nature and structure of pRi T-DNA genes and their use as a means for expression of a foreign structural gene in a plant genome. The remaining steps of the preferred embodiment for obtaining a genetically modified plant include inserting the pRi T_L-DNA TxCS/structural gene combination into T-DNA, transferring the modified T-DNA to a plant cell wherein the modified T-DNA becomes stably integrated as part of the plant cell genome, techniques for in vitro culture and eventual regeneration into whole plants, which may include steps for selecting and detecting transformed plant cells and steps of transferring the introduced gene from the originally transformed strain into commercially acceptable cultivars.

An advantage, which will be readily understood by those skilled in the art, of use of transcription controlling sequences disclosed herein

for controlling structural gene expression over previously published T-DNA TxCSs is that transcription of many pRI T-DNA ORFs is phenotypically and developmentally regulated (see Example 1.9). pT1 T-DNA genes are not known to be so regulated. Transcripts of ORFs 8, 11, 13, and 15 are more prevalent in roots than leaves, with the case of ORF 15 being particularly striking, while ORF 12 expression is specific to leaves and to a particular phenotype (T', see Example 1.9). Therefore, choice of a particular pRI T_L-DNA TxCS allows modulation of expression of a structural gene with which the TxCS is combined. For example, should one want expression of a structural gene to be much higher in roots than leaves; ORF15 provides the TxCS of choice.

A principal feature of the present invention in its preferred embodiment is the construction of T-DNA having an inserted foreign structural gene under control of a pRI T_L-DNA TxCS, i.e., between a promoter and a polyadenylation site, as these terms have been defined, supra, at least one of which is derived from pRI T_L-DNA. The structural gene must be inserted in correct position and orientation with respect to the desired pRI T_L-DNA promoter. Position has two aspects. The first relates to which side of the promoter the structural gene is inserted. It is known that the majority of promoters control initiation of transcription and translation in one direction only along the DNA. The region of DNA lying under promoter control is said to lie "downstream" or alternatively "behind" or "3' to" the promoter. Therefore, to be controlled by the promoter, the correct position of foreign structural gene insertion must be "downstream" from the promoter. The second aspect of position refers to the distance, in base pairs, between known functional elements of the promoter, for example the transcription initiation site, and the translational start site of the structural gene. Substantial variation appears to exist with regard to this distance, from promoter to promoter. Therefore, the structural requirements in this regard are best described in functional terms. As a first approximation, reasonable operability can be obtained when the distance between the promoter and the inserted foreign structural gene is similar to the distance between the promoter and the T-DNA gene it normally controls. Orientation refers to the directionality of the structural gene. That portion of a structural gene which ultimately codes for the amino terminus of the foreign protein is termed the

5' -end of the structural gene, while that end which codes for amino acids near the carboxyl end of the protein is termed the 3'-end of the structural gene. Correct orientation of the foreign structural gene is with the 5'-end thereof proximal to the promoter. An additional requirement in 5 the case of constructions leading to fusion protein expression is that the insertion of the foreign structural gene into the pRI T_L -DNA promoter-
donated structural gene sequence must be such that the coding sequences of the two genes are in the same reading frame phase, a structural requirement which is well understood in the art. An exception to this requirement exists in the case where an intron separates coding sequences derived 10 from a foreign structural gene from the coding sequences of the pRI T_L -DNA structural gene. In that case, both structural genes must be provided with compatible splice sites, and the intron splice sites must be so positioned that the correct reading frame for the pRI T_L -DNA promoter-
donated structural gene and the foreign structural gene are restored in 15 phase after the intron is removed by post-transcriptional processing. Differences in rates of expression or developmental control may be observed when a given foreign structural gene is inserted under control of different pRI T_L -DNA T_{CS}s. Rates of expression may also be greatly 20 influenced by the details of the resultant mRNA's secondary structure, especially stem-loop structures. Stability, ability to be excreted, intercellular localization, intracellular localization, solubility, target specificity, and other functional properties of the expressed protein itself may be observed in the case of fusion proteins depending upon the 25 insertion site, the length and properties of the segment of pRI T_L -DNA protein included within the fusion protein and mutual interactions between the components of the fusion protein that effect folded configuration thereof, all of which present numerous opportunities to manipulate and control the functional properties of the foreign protein product, depending 30 upon the desired physiological properties within the plant cell, plant tissue, and whole plant. Similarly to the promoter, the polyadenylation site must be located in correct position and orientation relative to the 3'-end of the coding sequence. Fusion proteins are also possible between the 3'-end of the foreign structural gene protein and a polypeptide 35 encoded by the DNA which serves as a source of the polyadenylation site.

A TxCS is comprised by two major functionalities: a promoter, which is absolutely necessary for gene expression, and a transcript terminator, being in the preferred embodiment a polyadenylation site, positioned respectively 5' and 3' to the structural gene. Although as exemplified herein these two portions of the TxCS are obtained from the same gene, this is not a requirement of the present invention. These 5' and 3' sequences may be obtained from diverse pRi T-DNA genes, especially pRi T_L -DNA genes, or one of these sequences may even be obtained from a non-pRi T-DNA gene. For instance, a promoter may be taken from a pRi T_L -DNA gene while the polyadenylation site may come from a plant gene.

In the Examples, a foreign structural gene is nested within a pRi T_L -DNA TxCS, suturing the structural gene into the TxCS at NdeI sites and placing the entire TxCS/structural gene combination between a pair of BamHI sites. As will be apparent to those of ordinary skill in the art, the TxCS/gene combination may be placed between any restriction sites convenient for removing the combination from the plasmid it is carried on and convenient for insertion into the plant transformation or shuttle vector of choice. Alternatives to the use of paired NdeI sites (5'...CATATG...3') at the ATG translational start include, but are not limited to, use of ClaI (5'...(not G)ATCGAT(G)...3') or NcoI (5'...CCATGG...3') sites. As will be understood by persons skilled in the art, other sites may be used for the promoter/structural gene suture as long as the sequence at the junction remains compatible with translational and transcriptional functions. An alternative to the suture of the promoter to the foreign structural gene at the ATG translational start is suturing at the transcriptional start or cap site. An advantage, especially for eukaryotic structural genes, of the use of this location is the secondary (stem-loop) structure of the foreign structural gene mRNA will not be disrupted thereby leading to an mRNA having translational activity more nearly resembling the activity observed in the organism which was the source of the gene. The restriction sites at the 5'- and 3'-ends of the structural gene need not be compatible. Use of cut sites cut by two different restriction enzymes at the two TxCS/structural gene junctions will automatically correctly orient the structural gene when it is inserted between the TxCS elements, though use of an extra restriction enzyme may necessitate removal of an additional set of inconvenient

restriction sites within the TxCS and the structural gene. The use of a single restriction enzyme to link both a promoter and a polyadenylation site to a particular structural gene is not required. Convenient sites within the pRi_L-DNA structural gene and 3' to the translational stop of the foreign structural gene may be used. When these sites have incompatible ends, they may be converted to blunt-ends by methods well known in the art and blunt-end ligated together.

Location of the TxCS/foreign structural gene combination insertion site within T-DNA or a T-DNA-derived vector is not critical as long as the transfer function of the T-DNA borders and any other necessary vector elements (e.g. a selectable or screenable marker) are not disrupted. The T-DNA into which the TxCS/structural gene combination is inserted may be obtained from any of the TIP plasmids, including both Ti and Ri plasmids. The TxCS/structural gene combination is inserted by standard techniques well known to those skilled in the art. The orientation of the inserted plant gene, with respect to the direction of transcription and translation of endogenous T-DNA or vector genes is not critical, either of the two possible orientations is functional. Differences in rates of expression might be observed when a given gene is inserted at different locations within T-DNA.

A convenient means for inserting a TxCS/foreign structural gene combination into T-DNA involves the use of a shuttle vector, as described in the Background. An Agrobacterium strain transformed by a shuttle vector is preferably grown under conditions which permit selection of a double-homologous recombination event which results in replacement of a pre-existing segment of a Ti or Ri plasmid with a segment of T-DNA of the shuttle vector. However, it should be noted that the present invention is not limited to the introduction of the TxCS/structural gene combination into T-DNA by a double homologous recombination mechanism; a homologous recombination event with a shuttle vector (perhaps have only a single continuous region of homology with the T-DNA) at a single site will also prove an effective means for inserting that combination into T-DNA as will insertion of a combination-carrying bacterial transposon.

An alternative to the shuttle vector strategy involves the use of plasmids comprising T-DNA or modified T-DNA, into which an TxCS/foreign

structural gene is inserted, said plasmids lacking vir genes and being capable of independent replication in an Agrobacterium strain. As reviewed in the Background, the T-DNA of such plasmids can be transferred from an Agrobacterium strain (e.g. A. rhizogenes, A. tumefaciens, or derivatives thereof) to a plant cell provided the Agrobacterium strain contains certain trans-acting vir genes whose function is to promote the transfer of T-DNA to a plant cell. Plasmids that contain T-DNA and are able to replicate independently in an Agrobacterium strain are herein termed "sub-TIP" plasmids. A spectrum of variations is possible in which the sub-TIP plasmids, which may be derived from Ri or Ti plasmids, differ in the amount of T-DNA contained. A "mini-TIP" plasmid retains all of the T-DNA from a TIP. "Micro-TIP" plasmids are deleted for all T-DNA but that surrounding the T-DNA borders, the remaining portions being the minimum necessary for the sub-TIP plasmid to be transferrable and integratable in the host cell. Sub-TIP plasmids are advantageous in that they are relatively small and relatively easy to manipulate directly, eliminating the need to transfer the gene to T-DNA from a shuttle vector by homologous recombination. After the desired structural gene has been inserted, they can easily be introduced directly into a Agrobacterium cell containing the trans-acting genes that promote T-DNA transfer. Introduction into an Agrobacterium strain is conveniently accomplished either by transformation of the Agrobacterium strain or by conjugal transfer from a donor bacterial cell, the techniques for which are well known to those of ordinary skill.

pRi T-DNA TxCS/structural gene combinations may be combined with pTi-derived Ti plasmids or sub-TIP vectors.

Modified T-DNA carrying a pRi T_L-DNA TxCS/structural gene combination can be transferred to plant cells by any technique known in the art (see Background). The resultant transformed cells must be selected or screened to distinguish them from untransformed cells. Selection is most readily accomplished by providing a selectable marker known to the art incorporated into the T-DNA in addition to the TxCS/foreign structural gene combination. Indeed, a pRi T_L-DNA TxCS can be a component of such a marker. In addition, the T-DNA provides endogenous markers such as the gene or genes controlling hormone-independent growth of Ti-induced tumors

in culture, the gene or genes controlling abnormal morphology of Ri-induced tumor roots, and genes that control resistance to toxic compounds such as amino acid analogs, such resistance being provided by an opine synthase (e.g. ocs). Screening methods well known to those skilled in the art include assays for opine production, specific hybridization to characteristic RNA or T-DNA sequences, or immunological assays. Additionally the phenotype of expressed foreign gene can be used to identify transformed plant tissue (e.g. insecticidal properties of the crystal protein).

Although the preferred embodiment of this invention uses a T-DNA-based Agrobacterium-mediated system for incorporation of the TxCS/foreign structural gene combination into the genome of the plant which is to be transformed, other means for transferring and incorporating the gene are also included within the scope of this invention. Other means for the stable incorporation of the combination into a plant genome additionally include, but are not limited to, use of vectors based upon viral genomes (e.g. see N. Brisson et al. (1984) *Nature* 310:511-514), minichromosomes, transposons, and homologous or nonhomologous recombination into plant chromosomes. Alternate forms of delivery of these vectors into a plant cell additionally include, but are not limited to, direct uptake of nucleic acid (e.g. see J. Paszkowski et al. (1984) *EMBO J.* 3:2717-2722), fusion with vector-containing liposomes or bacterial spheroplasts, micro-injection, and encapsidation in viral coat protein followed by an infection-like process. After introduction into a plant cell of a pRi T_L-DNA TxCS/structural gene combination, the combination will be contained by a plant cell. Furthermore, the combination will be flanked by plant DNA, unless utilizing a nonintegrating vector, e.g. a virus or minichromosome.

Regeneration of transformed cells and tissues is accomplished by resort to known techniques. An object of the regeneration step is to obtain a whole plant that grows and reproduces normally but which retains integrated T-DNA. The techniques of regeneration vary somewhat according to principles known in the art, depending upon the origin of the T-DNA, the nature of any modifications thereto and the species of the transformed plant. In many plant species, cells transformed by pRi-type T-DNA are readily regenerated, using techniques well known to those of ordinary skill, without undue experimentation. Plant cells transformed by pTi-type

T-DNA can be regenerated, in some instances, by the proper manipulation of hormone levels in culture. Preferably, however, the Ti-transformed tissue is most easily regenerated if the T-DNA has been mutated in one or both of the tmr and tms genes. It is important to note that if the mutations in tmr and tms are introduced into T-DNA by double homologous recombination with a shuttle vector, the incorporation of the mutation must be selected in a different manner than the incorporation of the TxCS/structural gene combination; e.g. one might select for tmr and tms inactivation by chloramphenicol resistance while one might select for TxCS/foreign gene integration by kanamycin resistance. The inactivation of the tms and tmr loci 5 may be accomplished by an insertion, deletion, or substitution of one or more nucleotides within the coding regions or promoters of these genes, the mutation being designed to inactivate the promoter or disrupt the structure of the encoded proteins (e.g. the T-DNA of NRRL B-15821, or the 10 pTi of A3004, L. W. Ream et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80:1660-1664). Resultant transformed cells are able to regenerate plants 15 which carry integrated T-DNA and express T-DNA genes, such as an opine synthase, and also express an inserted pRI T_L-DNA TxCS/structural gene combination. These serve as parental plant material for normal progeny 20 plants carrying and expressing the pRI T_L-DNA TxCS/heterologous foreign structural gene combination, and for seeds containing the combination, in the preferred embodiments the combination being integrated into a plant chromosome and flanked by plant DNA.

25 The genotype of the plant tissue transformed is often chosen for the ease with which its cells can be grown and regenerated in in vitro culture and for susceptibility to the selective agent to be used. Should a cultivar of agronomic interest be unsuitable for these manipulations, a more 30 amenable variety is first transformed. After regeneration, the newly introduced TxCS/foreign structural gene combination is readily transferred to the desired agronomic cultivar by techniques well known to those skilled in the arts of plant breeding and plant genetics. Sexual crosses 35 of transformed plants with the agronomic cultivars yielded initial hybrid. These hybrids can then be back-crossed with plants of the desired genetic background. Progeny are continuously screened and selected for the continued presence of integrated T-DNA or for the new phenotype resulting from expression of the inserted foreign gene. In this manner, after

5

a number of rounds of back-crossing and selection, plants can be produced having a genotype essentially identical to the agronomically desired parents with the addition of a inserted pRi T-DNA promoter/foreign structural gene combination or of a foreign structural gene/polyadenylation site combination.

10

EXAMPLES

The following Examples are presented for the purpose of illustrating specific embodiments within the scope of the present invention without limiting the scope; the scope being defined by the Claims. Numerous variations will be readily apparent to those of ordinary skill in the art.

15

20

25

These Examples utilize many techniques well known and accessible to those skilled in the arts of molecular biology and manipulation of TIPs and Agrobacterium; such methods are fully described in one or more of the cited references if not described in detail herein. Enzymes are obtained from commercial sources and are used according to the vendor's recommendations or other variations known to the art. Reagents, buffers and culture conditions are also known to those in the art. Reference works containing such standard techniques include the following: R. Wu, ed. (1979) Meth. Enzymol. 68, R. Wu et al., eds. (1983) Meth. Enzymol. 100 and 101, L. Grossman and K. Moldave, eds. (1980) Meth. Enzymol. 65, J. H. Miller (1972) Experiments in Molecular Genetics, R. Davis et al. (1980) Advanced Bacterial Genetics, R. F. Schleif and P. C. Wensink (1982) Practical Methods in Molecular Biology, and T. Maniatis et al. (1982) Molecular Cloning. Additionally, R. F. Lathe et al. (1983) Genet. Engin. 4:1-56, make useful comments on DNA manipulations.

30

35

Textual use of the name of a restriction endonuclease in isolation, e.g. "BclI", refers to use of that enzyme in an enzymatic digestion, except in a diagram where it can refer to the site of a sequence susceptible to action of that enzyme, e.g. a restriction site. In the text, restriction sites are indicated by the additional use of the word "site", e.g. "BclI site". The additional use of the word "fragment", e.g. "BclI fragment", indicates a linear double-stranded DNA molecule having ends generated by action of the named enzyme (e.g. a restriction fragment). A

phrase such as "BclI/SmaI fragment" indicates that the restriction fragment was generated by the action of two different enzymes, here BclI and SmaI, the two ends resulting from the action of different enzymes. Note that the ends will have the characteristics of being "sticky" (i.e. having 5 a single-stranded protrusion capable of base-pairing with a complementary single-stranded oligonucleotide) or "blunt" and that the sequence of a sticky-end will be determined by the specificity of the enzyme which produces it.

10 In the Examples and Tables, the underlining of a particular nucleotide in a primer or other sequence indicates the nucleotide which differs from the naturally found sequence, being an insertion or substitution of one or more nucleotides. The use of lower case for two adjacent nucleotides brackets one or more nucleotides that have been deleted from the 15 native sequence. Unless otherwise noted, all oligonucleotide primers are phosphorylated at their 5'-ends, are represented 5'-to-3', and are synthesized and used as referenced in Example 5.

Plasmids are usually prefaced with a "p", e.g., pRiA4 or p8.8, and strain parenthetically indicate a plasmid harbored within, e.g., 20 A. rhizogenes (pRiA4) or E. coli HB101 (p8.8). Self-replicating DNA molecules derived from the bacteriophage M13 are prefaced by an "m", e.g. mWB2341, and may be in either single-stranded or double-strand form. A. tumefaciens (pTi15955) is on deposit in ATCC 15955, E. coli C600 (pRK-203-Kan-103-Lec) as NRRL B-15821, E. coli HB101 (pLJ40) as NRRL 25 B-15957, and E. coli HB101 (EcoRI e36) as NRRL B-15958 (as deposited EcoRI e36 was designated EcoRI 3a); other deposited strains are listed in column 3 of Table 7.

30 The DNA constructions described in these Examples have been designed to enable any one of the eukaryotic TxCSs of pRi T_L-DNA to be combined with any of four foreign structural genes. Towards that end, the structural genes, the TxCSs, and the TxCS/structural gene combinations have been placed on DNA "cassettes", having the properties that, after initial 35 modifications have been made, any structural gene may be readily inserted into any TxCS without further modification, and any TxCS/structural gene combination may be isolated by a simple procedure applicable to all such combinations. All combinations are thereby equivalent when being inserted

into the plant transformation vector of choice. The initial modifications of the TxCSs are all analogous to each other and the initial modifications of the structural genes are also all analogous to each other. These Examples often involve the use of a common strategy for multiple constructions that differ only in items such as choice of restriction enzymes, DNA fragment size, ORFs encoded, plasmids generated or used as starting material, specific numbers and sequences of oligonucleotides used for mutagenesis, sources of plasmids, and enzyme reactions utilized. For the sake of brevity, the DNA manipulations and constructions are generally described once, the differing items being detailed by reference to a particular column in a particular Table, a particular series of manipulations used in a particular construction occupying horizontal lines within that Table. One combination, the ORF 11 TxCS with the crystal protein structural gene, is also detailed in the text.

The following is an outline, diagrammed schematically in Figure 3, of a preferred strategy used to make the exemplified DNA constructions detailed in Examples 3 through 6. Endogenous NdeI sites are removed from the M13-based vector mWB2341, resulting in a vector designated mWB2341(Nde) (Example 3.1). Large fragments of T-DNA are introduced into mWB2341(Nde) in a manner that also eliminates the vector's BamHI site (Example 3.2). Endogenous T-DNA NdeI and BamHI sites are then removed (Example 3.3) and novel sites are introduced. NdeI sites are introduced at and near the translational start and stop sites, respectively, so that a foreign structural gene on a NdeI fragment may replace the endogenous ORF structural gene. BamHI sites are introduced approximately 0.3 kbp 5' to and 3' from the transcriptional start and stop signals, respectively, so that the TxCS/structural gene combination eventually constructed may be removed on a BamHI fragment (Example 3.4). The structural genes, which fortuitously have no internal NdeI or BamHI sites, are introduced into mWB2341 (Nde) (Example 4.1) and NdeI sites are introduced at and after the translational start and stop sites (Examples 4.2 and 4.3). The structural genes are removed from their vectors on "DNA cassettes" by digestion with NdeI and are inserted into any desired TxCS which has had its endogenous structural gene removed by NdeI digestion (Example 6.1). The TxCS/foreign structural gene combinations are then removed from their vector by digestion with BamHI and inserted into the plant transformation vectors of

choice (Example 6.2). It is recognized that construction strategies utilizing fortuitously located restriction sites might be designed by persons of ordinary skill which might be simpler for some particular TxCS/structural gene combination than the generalized DNA cassette strategy utilized herein; however, DNA cassettes are a better approach when one is trying to achieve flexibility in the choice and matching of many diverse TxCSs and structural genes.

Example 1

This Example provides disclosure, analysis, and discussion of the pRi T_L -DNA sequencing results.

1.1 Summary of results

pRi T_L -DNA was sequenced and eighteen open reading frames (ORFs), two of which (7 and 18) were clearly prokaryotic in nature, were found. Eleven ORFs had canonical eukaryotic promoter and polyadenylation elements (ORFs 1, 2, 3, 6, 8, 11, 12, 13, 14, 15 and 16). These ORFs were distributed within an about 19.4 kilobase pair (kbp) segment of pRi T_L -DNA integrated into the genome of C. arvensis clone 7. DNA encoding ORFs 8, 11, 12, 13, and 15 was observed to be transcribed in tobacco.

1.2 Sequence of pRi T_L -DNA

A physical map of the pRi T_L -DNA region is shown in Figure 1 along with pRi subclones and the nucleotide sequencing strategy used. Nine-tenths of the sequence obtained was determined from both DNA strands, the remaining tenth being sequenced more than once from the same DNA strand. A nucleotide sequence of 21,126 base pairs (bp) was obtained, which included a 19.4 kbp pRi T_L -DNA region identified in the genome of C. arvensis clone 7, and is presented in Figure 2, 5'-to-3' corresponding to left-to-right as mapped in Fig. 1. DNA was sequenced from the 5'-end of BamHI fragment 32 to about 2216 bp into EcoRI fragment 3b (3'-end) (see Fig. 1). The cleavage sites for over seventy restriction enzymes were determined; cleavage positions for enzymes with less than nineteen sites are listed in Table 1.

1.3 T_L -DNA border repeats

Genomic hybridization and DNA sequence analyses of the T_L -DNA region integrated into the genome of C. arvensis clone 7 showed the exact location of a left plant/T-DNA junction and an approximate position for a right pRi T_L -DNA/plant junction (F. Leach (1983) Ph.D. Thesis, Universite de Paris-Sud, Centre d'Orsay, France). The left plant DNA/T-DNA junction was between position 570 and 571, as defined in Fig. 2. The left 25 bp T-DNA border repeat sequence was located between positions 520 and 544. The right boundary of T_L -DNA of RiA4-transformed C. arvensis could vary over a 8 kbp region. The complete 21,126 bp of pRi T_L -DNA region was scanned for the presence of a 25 bp consensus sequence derived by comparison with published sequences, 5' TGGCAGGATATAT ^{TCACT} AGGAG GCTAA ^{ATC} 3'. Twenty-seven nucleotide sequences matching this consensus at 15 or more bases were identified. Included among these sequences were the 25 bp nucleotide sequences starting (5') at positions 520 (matching at 23 of 25 bases), and 19,966 (17 of 25) (see Fig. 2). These two positions were near the T-DNA/plant junctions of a transformed Nicotiana glauca tissue (F. F. White et al. (1983) Nature 301:348-350) and C. arvensis clone 7, as determined by comparison of genomic restriction maps of transformed plant DNA and pRiA4 DNA. Other matches were found at positions 154, 576, 725, 3244, 6316, 6365, 7209, 7379, 8697, 10339, 10436, 11079, 11232, 12313, 13832, 14235, 14510, 15145, 16285, 17071, 17483, 18121, 18273, 18368, and 18797. The eleven previously published 25 bp border repeat sequences were as little as 64% homologous to each other, thus indicating that many of these pRi border sequences could be functional. Genomic hybridization analysis of the pRi T_L -DNA region in tobacco (D. Tepfer (1984) Cell 37:959-967) showed a much smaller T_L -DNA with the left junction probably involving a border sequence at either position 6316 or 6365.

1.4 Identification of open reading frames

Analysis of the nucleotide sequence presented in Fig. 2 revealed the presence of sixteen ORFs starting with an ATG initiation codon and extending over 300 nucleotides. The locations, sizes, and molecular weights of the putative translational polypeptides of these ORFs are listed in Table 3. Two additional ORFs (9 and 10) were shorter than 300 nucleotides but were included in Table 3 because they satisfied other criteria (see

below). The size of the ORFs ranged from 255 nucleotides (ORF 9) up to 2280 nucleotides (ORF 8), encoding polypeptides ranging in size from 9600 to 85,000 daltons, respectively. However, the actual size of an RNA transcript encoding an ORF could be considerably larger than that listed in 5 Table 3 because 5' and 3' noncoding regions and 3'-polyadenylic acid tails were not included.

Though to date no introns have been found in any of the fourteen sequenced pTi T-DNA genes, (R. F. Barker *et al.* (1983) *Plant Mol. Biol.* 2:335-350), J. Gielen *et al.* (1984) *EMBO J.* 3:835-846), introns are 10 present in some plant nuclear genes; pRi T_L -DNA genes could have introns. Transcript mapping (Example 1.9) did not generally indicate spliced mRNA. However, analysis of mRNA encoded between positions 6500 and 9000 detected two transcripts, a 2300 base transcript as predicted for ORF 8 and an unpredicted 650 base transcript. The nucleotide sequence of 15 the only other ORF in this region, ORF 9, suggested a transcript of about 450 bases, about half the size as found. The coding region of ORF 8 was scanned for sequences which matched consensus donor (5' exon...TG*GT^A_GAGT...intron3', the "*" indicating the splice site) and acceptor (intron...AAATTG^ATAG*G^G...exon) intron splice sequences and conformed to the G-T/A-G rule (R. Breathnach *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:4853-4857) and a plant consensus sequence (J. L. Slightom 20 *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1897-1901). Splicing between an acceptor at position 8943 and a donor at positions 7283, 7327, 7374, 25 7701, or 7894 would result in a second transcript having a translation initiation codon-polyadenylation site distance of 724, 758, 943, 1270, or 1325 bp, respectively, which is in the size range observed. Proper processing of an intron-containing genes in T-DNA has been observed (e.g. N. Murai *et al.* (1983) *Science* 222:476-482).

No homology greater than random was found to exist in coding or 30 noncoding sequences between pRi T_L -DNA and octopine pTi T-DNA (Barker *et al.*, supra), consistent with the lack of cross-hybridization between pRi T_L -DNA and octopine pTi T-DNA observed by G. A. Huffman *et al.* (1984) *J. Bacteriol.* 157:269-276, and L. Jouanin (1984) *Plasmid* 12:91-102.

1.5 Translational initiation codons

Eukaryotic translation is preferentially initiated at the first AUG of an mRNA; and A or G at position -3 and G at position +4 may facilitate recognition of functional AUG codons. This $\text{G}^{\text{A}}\text{XXAUGG}$ consensus is referred to as the ribosome binding site (M. Kozak (1981) *Nucl. Acids Res.* 9:5233-5252; M. Kozak (1983) *Cell* 34:971-978). The number of amino acids and calculated molecular weights for the putative pTi T_L -DNA protein products (Table 3) were derived by assigning the first in-phase AUG codon as the initiator codon. The art has not ruled out use of secondary AUG codons as translation initiation codons (M. Kozak (1983) *Microbiol. Rev.* 47:1-45).

Initiator codon DNA sequences are listed in Table 3 below the consensus eukaryotic ribosome binding site. Eight of the eighteen ORFs had first AUG codons which conform with this consensus sequence (ORFs 1, 7, 8, 10, 11, 12, 14, and 18). Of the ten remaining ORFs, four had downstream, 15 in-phase AUG codons which conformed with the consensus sequence: ORF 2, 287 bp downstream; ORF 3, 160 bp; ORF 6, 344 bp; ORF 13, 203 bp; and ORF 17, 105 bp (see Fig. 2). The remaining six ORFs (2, 4, 5, 9, 15, and 16) did not have any AUG codons which conform to the consensus sequence followed by 300 bp in-phase ORFs. The presence of a consensus ribosome binding AUG codon is not necessary for translation initiation of T -DNA mRNAs; four abundantly transcribed octopine pTi T_L -DNA genes are initiated at AUG codons which do not conform to the consensus sequences.

Several pTi T -DNA ORFs are actively transcribed in *E. coli* mini-cells (G. Schröder *et al.* (1983) *EMBO J.* 2:403-409). Translational initiation in *E. coli* and most prokaryotes generally start at an AUG codon that is proceeded by a G-rich ribosome binding site (J. Shine and L. Dalgarno (1974) *Proc. Natl. Acad. Sci. USA* 71:1342-1346). Sequences which may function as prokaryotic ribosome binding sites were observed ahead of the pTi T_L -DNA ORF 4, 5, 7, 9, and 18 initiation codons.

30

1.6 Codon usage

Most pTi T_L -DNA ORFs were observed to fit pTi T_L -DNA codon preference patterns, thereby indicating that they are functional after integration into a plant genome, notable exceptions being ORFs 7 and 18.

1.7 Locations of transcription controlling sequences

Comparisons of nucleotide sequences from the 5'-flanking regions of many eukaryotic genes have revealed consensus locations and sequences of several DNA elements which may be important in regulating RNA polymerase II-mediated transcription (S. L. McKnight and R. Kingsbury (1982) *Science* 217:316-324). These characteristic eukaryotic promoter elements are the "TATA-element", located 25-30 bp upstream (5') from the start of transcription, and the "CCAAT-element", located 40-50 nucleotides upstream from the TATA-element (C. Benoist *et al.* (1980) *Nucl. Acids. Res.* 8:127-142; A. Efstratiades *et al.* (1980) *Cell* 21:653-668). Similar promoter elements have been found in the 5'-flanking regions of many plant and pTi-T-DNA genes; pTi15955 T-DNA (Barker *et al.*, *supra*) and pTiAch5 T_L-DNA (Gielen *et al.*, *supra*) have sequences resembling these TATA and CCAAT promoter elements located in the 5'-flanking regions of eight T_L-DNA and six T_R-DNA ORFs (i.e. have "eukaryotic-looking" promoters) All eight eukaryotic-looking pTi T_L-DNA ORFs are transcribed and at least five of six eukaryotic-looking pTi T_R-DNA ORFs are known to be transcribed.

The presence of TATA and CCAAT promoter elements in 5'-flanking regions of pTi T_L-DNA ORFs indicated that a particular ORF was part of a functional gene. Most pTi T_L-DNA ORFs (16 of 18) were flanked by sequences (Table 3) that closely resembled these eukaryotic promoter elements. The amount of sequence identity between the promoter elements and the consensus sequences was very high; ORFs 2 and 12 had promoter elements which matched the consensus sequences while the promoter elements from the other thirteen ORFs did not vary by more than three mismatches. These results were consistent with the degree of homology found for promoter elements from pTi T-DNA ORFs (Barker *et al.*, *supra*; Gielen *et al.*, *supra*).

pTi T_L-DNA open reading frames 1, 4, 8, 10, 13, 14, and 17 were flanked by multiple promoter elements. ORFs 7 and 18 were not flanked by sequences resembling eukaryotic promoter elements and were not expected to be transcribed in plant tissues. ORFs 4, 5, 7, and 9 overlapped ORFs 5, 6, and 8 on the opposite strand (Fig. 1, Table 2); the larger ORFs (5, 6, and 8) were more likely to be transcribed because DNA encoding overlapping, antiparallel ORFs in pTi T-DNA was found to be transcribed from either one strand or the other (Gielen *et al.*, *supra*).

Comparison of polyadenylation sites present in the 3'-noncoding regions of plant genes indicates a preference for the hexanucleotide, AATAAA (J. Messing *et al.* (1983) in Genetic Engineering of Plants, ed.: A. Hollaender, pp. 211-227), however, variations have been observed for plant genes, e.g. AATAAG and GATAAA. Many pTi T-DNA ORFs are also followed by AATAAA sequences. The remaining pTi T-DNA ORFs are followed by polyadenylation sites which vary only slightly, e.g. AATAAT, TATAAA, or AATGAA; AATAAT is known to function for the ocs gene (H. DeGreve *et al.* (1982) *J. Mol. Appl. Genet.* 499-511).

10 Presumptive pRI T_L-DNA polyadenylation sites and their locations are
 listed in Table 3. Ten ORFs (2, 4, 6, 8, 9, 11, 12, 13, 14, and 15) had
 the consensus hexanucleotide, AATAAA, near their 3'-ends, whereas only two
 (ORFs 7 and 18) did not contain any related sequence (Table 3, Fig. 2).
 The remaining ORFs (1, 3, 10, and 16) had polyadenylation sites closely
 15 related to those described above. ORFs 8, 10, 12, 13, and 14 were
 followed by multiple polyadenylation signals. Multiple polyadenylation
 sites have also been observed in several pTi T-DNA genes (P. Dhaese *et al.*
 (1983) EMBO J. 2:419-426; Gielen *et al.*, *supra*).

1.8 ORF locations with respect to base composition

The G+C content of the large Agrobacterium plasmids is about 59% (S. Sheikholeslam *et al.* (1979) *Phytopathol.* **69**:54-58). In contrast, pRi T_L-DNA had very A+T-rich regions flanking the eukaryotic ORFs while coding regions had G+C contents in the range of 50%. Plant genes can also have A+T-rich flanking sequences.

1.9 Detection of transcripts

The T_L -DNA left junction with plant DNA found in an A. rhizogenes transformed tobacco tissue, clone 9, was between the position 6361 HindIII site and the position 7585 EcoRI site, while the right border was to the right of the position 19,918 KpnI site (see Example 1.3). Hybridization of nick-translated pRi T_L -DNA probes to membrane filter-bound replicas of the gels ("Northern blots") clearly showed transcripts carrying ORFs 8 and 13. An observed transcript of about 950 nucleotides which hybridized with pRi T_L -DNA between EcoRI sites at positions 9077 and 13,445 was assigned to ORF 11. An observed transcript of about 1400 nucleotides which hybri-

dized with sequences spanning the position 17,059 EcoRI site was assigned to ORF 15. An observed transcript of about 800 nucleotides which hybridized with pRI T_L -DNA between the positions 9077 and 13,445 EcoRI sites was assigned to ORF 12.

5 The relative abundances of pRI T_L -DNA transcripts in clone 9-derived plants were observed to be a function of organ (leaves vs. roots) and phenotype (T vs. T'; see Tepfer (1984) supra). With the exception of the transcript corresponding to ORF 12, pRI T_L -DNA transcripts were more prevalent in roots than in leaves, with a particularly striking case being 10 the mRNA assigned to ORF 15. Expression of the transcript assigned to ORF 12 was leaf specific and was correlated with the T' phenotype.

RNA from C. arvensis tissue transformed by pRI T_L -DNA which included sequences encoding ORFs 1-6 also hybridized with pRI T_L -DNA.

15 1.10 Conclusions

The data discussed above (Examples 1.2, 1.4-1.8) indicated that of the ORFs flanked by eukaryotic transcription controlling sequences (ORFs 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17), ORFs 1, 2, 3, 6, 8, 11, 12, 13, 14, 15, and 16 were most likely to be transcribed. 20 In tobacco tissue transformed by DNA encoding ORFs 8-18, transcription of DNA region encoding ORFs 8, 11, 12, 13, and 15 has been detected (Example 1.9).

Example 2

25 This Example discloses materials and methods used to obtain the results disclosed in Example 1.

2.1. Materials

30 Restriction endonucleases AvaI, BamHI, BglIII, EcoRI, HindIII, KpnI, PstI, PvuII, Sall, StuI, XbaI, and XhoI were obtained from Promega-Biotec. Enzymes AccI, Clal, DraI, MstI, MstII, NarI, NcoI, XmnI, and XorII were obtained from New England Biolabs. Polynucleotide kinase was from P-L Biochemicals and bovine alkaline phosphatase was from Boehringer-Mannheim. [γ -³²P] ATP (2000-3000 Ci/mole) was obtained from New England Nuclear. Chemicals used for DNA sequencing were obtained from the vendors recommended by A. M. Maxam and W. Gilbert (1980) *Meth. Enzymol.* 65:499-

560. X-ray film on rolls (20 cm x 25 m) XAR-351 was obtained from Kodak. DuPont Quanta III intensifying screens (35 cm x 1 m) were cut in half to fit sequencing gels (17.5 cm x 1 m). DNA sequencing gel stands, designed for gels measuring 20 cm x 104 cm, and safety cabinets were from Fotodyne Inc., New Berlin, Wisconsin. Water jacket thermostating plates were constructed using 1/4 inch thick plate glass glued together by 100% silicone rubber.

2.2 DNA isolation

10 Procedures for the isolation and mapping of plasmid and cosmid sub-clones of the closely-related R_i plasmids pRiA4 and pRiHRI have been published: A4 subclones: EcoRI e36 (EcoRI 3a), BamHI 8a, e16 (contains R_i EcoRI fragments 15, 36, and 37a) by F. Leach (1983) Ph.D. Thesis, Universite de Paris-Sud, Centre d'Orsay; and pRiHRI subclones: pLJ40 (i.e. cosmid 40) and EcoRI 3b by L. Jouanin (1984) Plasmid 12:81-102. 15 Plasmid DNAs were prepared as described by H. C. Birnboim and J. Doly (1979) Nucl. Acids Res. 7:1513-1523, followed by two CsCl, ethidium bromide gradient bandings.

2.3 DNA sequencing

20 DNA sequences were determined using the chemical method, essentially as described by Maxam and Gilbert, supra. Generally, 10-20 μ g of plasmid DNA was digested with the appropriate restriction enzyme, followed by removal of the 5' terminal phosphate with 2-3 units of calf intestinal 25 alkaline phosphatase. Reactions were done in 100 mM Tris pH 8.4, 55°C for 30 min. Both restriction enzyme and phosphatase were removed by two phenol and one chloroform extractions. DNA samples were then precipitated with ethanol, desalting with 70% ethanol, dried, and then resuspended in 15 μ l denaturation buffer (50 mM Tris-HCl (pH 9.5), 5 mM spermidine, and 30 0.5 mM EDTA) and 15 μ l H₂O. End-labeling with [γ -³²P]ATP and isolation of end-labeled fragments were as described by Maxam and Gilbert, supra. Care was taken to avoid sequencing errors resulting from the presence of hydrazine-unreactive 5-methylcytosine bases, found after growth in E. coli at 35 the second cytosine base of EcoRII or BstNI restriction enzyme sites (J. L. Slightom et al. (1980) Cell 21:627-638).

Conditions for chemical reactions, at 20°C, were as follows: 1 μ l dimethyl sulfate for G, 30 sec.; 30 μ l of formic acid 95% for A, 2.5 min.; 30 μ l of hydrazine 95% for C+T and C, 2.5 min. DNA samples were electrophoresed 14 hours, at 2500 V at constant voltage on gels 20 cm wide, 5 104 cm long and 0.2 mm thick. Constant gel temperatures (50°C) were maintained using a water-jacketed plate on one side of the gel sandwich. The opposite plate of the sandwich was treated with γ -methacryloxypropyl-trimethoxy silane (Sigma 6514) as described by H. Garoff and W. Ansorge 10 (1980) *Analyst. Biochem.* 115:450-457, to bind the acrylamide chemically to the glass. Gel pouring, loading, and autoradiography have been described by R. F. Barker et al. (1983) *Plant Mol. Biol.* 2:335-350, and J. L. Slightom et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:1897-1901.

Computer programs for DNA sequence analysis were supplied by the University of Wisconsin Genetics Computer Group.

15

Example 3

This Example teaches the manipulation of pRi T_L-DNA TxCSs preparatory to insertion of a foreign structural gene.

3.1 Removal of NdeI sites from an M13-based vector

These Examples extensively use oligonucleotide-directed, site-specific mutagenesis of DNA (see Example 5.2). Although individuals skilled in the art may choose to use double-stranded DNA methods for such mutagenesis, as exemplified herein single-stranded methods are used. In 25 general, single-stranded methods utilize M13-based vectors having inserted E. coli lac gene sequences. Wild-type M13 contains three NdeI sites while the lac sequences contain no NdeI site; BamHI sites are absent from both M13 and lac. Removal of these NdeI sites, described below, by site-specific mutagenesis may prove essential when replacing a T-DNA structural 30 gene with a heterologous foreign structural gene (Example 6.1). M13-based vectors include mWB2341 and related vectors (W. M. Barnes et al. (1983) *Meth. Enzymol.* 101:98-122; W. M. Barnes and M. Bevan (1983) *Nucl. Acids Res.* 11:349-368), and the M13mp-series of vectors (e.g. see J. Norrander et al. (1983) *Gene* 26:101-106, J. Messing and J. Vieira (1982) *Gene* 19:269-276). mWB2341 and related vectors are linearized by digestion with EcoRI and HindIII and the resultant sticky-ends are converted to blunt-

ends by incubation with the Klenow fragment of E. coli DNA polymerase I. Most of the M13mp-series vectors can be linearized by at least one blunt-end-forming restriction endonuclease (e.g. SmaI or HincII). In the alternative, particular single-stranded DNA vectors may be preferred for some operations; other vectors may be substituted for those referred to above with minor modification of procedures described herein, as will be understood by those of ordinary skill in the art. Also in the alternative, double-stranded DNA vectors might be substituted (see references cited in Example 5.2).

Single-stranded DNA (ssDNA) of the viral form of an M13-based vector is isolated and subjected to oligonucleotide-directed site-specific mutagenesis, described in detail in Examples 3.3 and 5, after hybridization to 5'CAATAGAAAATTCATAGGGTTTACC3', 5'CCTGTTAGTATCATAGCGTTATACC3', and 5'CATGTCAATCATTTGTACCCCCGGTTG3', thereby removing three NdeI sites which will later prove to be inconvenient without changing the translational properties of the encoded proteins. A mutated M13-based vector lacking three NdeI sites is identified and designated m13(Nde).

3.2 Subcloning pRi T₁-DNA into an M13-based vector

DNA of a plasmid listed in Table 4, column 1 (e.g. pLJ40 for manipulations of the ORFs 11, 12, and 13 promoters and polyadenylation sites) (see Example 2.2 for the sources of these plasmids) is isolated and digested to completion with the restriction enzyme(s) listed in Table 4, column 2 (e.g. SmaI and MstII for ORFs 11, 12, and 13). DNAs of e36 and pLJ40 are respectively harbored by the deposited strains NRRL B-15958 and NRRL B-15957. (Alternatively, pRiA4 DNA or pRiHRI DNA may be isolated and digested with the enzyme(s) listed in Table 4, column 2.) 5' or 3'-protruding-ends are then converted to blunt-ends by incubation with the Klenow fragment of E. coli DNA polymerase I or T4 DNA polymerase, respectively, and all four deoxynucleotide triphosphates. The resulting mixture of DNA fragments separated by agarose gel electrophoresis and a fragment whose size is listed in Table 4, column 3 (e.g. 5.2 kbp for ORFs 11, 12, and 13) is eluted from the gel.

Covalently-closed-circular DNA (cccDNA) of the replicative form (RF) of the M13-based vector m13(Nde) is isolated, converted to a linear, blunt-ended DNA, and has its 5'-phosphates removed by incubation with

phosphatase. The resulting linearized vector is purified by gel electrophoresis and is mixed with and ligated to the T-DNA fragment isolated above. After transformation of the resulting mixture into E. coli, viral DNAs and RFs are isolated from transformants and screened by restriction and hybridization analysis for the presence of inserts which when in single-stranded viral form, are complementary to the sequence as presented in Fig. 1 and which carry the complete DNA sequence of ORFs listed in Table 4, column 4. The virus which infects the selected colony is designated as listed in Table 4, column 5 (e.g. mR4 for ORFs 11, 12, and 13).

10

3.3 Removal of endogenous NdeI and BamHI sites from pRI T_L -DNA

A vector designated as listed in Table 5, column 1 (e.g. mR4' for manipulations of the ORFs 11, 12, and 13 promoters and polyadenylation sites) is prepared from the vector listed in the corresponding line of Table 5, column 2 (e.g. mR4 for ORFs 11, 12, and 13) by primer extension after hybridization to the oligonucleotides listed in Table 5, column 3 (e.g. 5'GATTAGATAGTCCAGATGAGCATGTGC3', 5'GCAAATCGGAGCCCCCTCGAATAGG3', 5'GCAATTGGGAGCCATTGTGATGTGAG3', and 5'CGGTTACGCGGAGGCCTATGCGGAGCGCC3' for ORFs 11, 12, and 13). This operation removes indigenous BamHI sites and NdeI sites, the sites designated in Table 5, column 4 being at pRI T_L -DNA positions listed in column 5 (e.g. for ORFs 11, 12, and 13, an NdeI site at position 10,305 and BamHI sites at positions 11,198, 11,278, and 12,816), which may be present which may prove inconvenient in later manipulations. (Note that there are no BamHI or NdeI sites in mR5.) The sites may be removed one at a time by hybridization of a particular oligonucleotide to the ssDNA viral form of the vector listed in Table 5, column 2 (e.g. mR4 for ORFs 11, 12, and 13), incubation of the primer/viral DNA complex with the Klenow fragment of E. coli DNA polymerase I, all four deoxynucleotide triphosphates, and DNA ligase, enrichment of resulting cccDNA molecules, transformation into E. coli selection of transformants, and isolation of RF followed by restriction enzyme analysis to identify a clone missing the undesired restriction sites. These steps are repeated for each site which is to be removed. Alternatively, the vector listed in Table 5, column 2 (e.g. mR4 for ORFs 11, 12, and 13) may be simultaneously hybridized to all of the oligonucleotides listed in Table 5, column 3 and then carried through the mutagenesis pro-

cedure thereby attempting, the procedure not being 100% efficient, to eliminate all of the sites in a single operation.

3.4 Placement of novel NdeI and BamHI sites in pRI T_L-DNA

A vector designated as listed in Table 6, column 1 (e.g. mORF 11 for manipulations of the ORF 11 promoter and polyadenylation site) is prepared from the vector listed in the corresponding line of Table 5, column 2 (e.g. mR4' for ORF 11) by primer extention after hybridization to the oligonucleotides listed in Table 6, column 3 (e.g.

5 5'GCTGCGAAGGGATCCCTTGTGCGCC3', 5'CGCAAGCTACAACATCATATGGGGCGG3',
10 5'GGGATCCATATGTGATGTGAGTTGG3', 5'GCCTAAGAAGGAATGGTGGATCCATGTACGTGC3' for
15 ORF 11) as described above and in Example 5. This has the effect of
introducing NdeI sites (5'...CATATG...3') at the translational start site
(ATG) and near the translational stop site (TAA, TGA, or TAG), and of
20 introducing BamHI sites (5'...GGATCC...3') in the sequences flanking the
T-DNA gene, usually approximately 0.3 kbp from the transcriptional start
and polyadenylation sites. The first and fourth oligonucleotide of each
25 quartet listed in Table 6, column 3 introduces BamHI sites while the
second and thirds introduce NdeI sites. These sites are located in the
corresponding pRI T_L-DNA at the approximate position listed in Table 6,
30 column 4. For example, for manipulation of ORF 11,
35 5'GCTGCGAAGGGATCCCTTGTGCGCC3' and 5'GCCTAAGAAGGAATGGTGGATCCATGTACGTGC3'
introduce BamHI sites and position 9,974 and 12,001, respectively, while
5'CGCAAGCTACAACATCATATGGGGCGG3' and 5'GGGATCCATATGTGATGTGAGTTGG3' intro-
duce NdeI sites at positions 10,679 and 11,286, respectively. The size
and locations of the TxCS-carrying DNA segments used in these Examples may
be calculated from the positions listed in Table 6, column 4 and the
orientations defined in Table 2 and Fig. 1. Positions listed in Table 6,
50 column 4, of pairs of NdeI and BamHI sites define promoter-bearing (P) and
polyadenylation site-bearing (A) DNA segments as indicated by "P"s and
"A"s, respectively, in column 5, the segments having approximate sizes as
indicated in column 6. For example, the ORF 11 promoter is on an approxi-
mately 715 bp DNA segment located between artificial NdeI and BamHI sites
at approximate positions 11,286 and 12,001, respectively, while the ORF 11
35 polyadenylation sites is on an approximately 705 bp DNA segment located
between artificial BamHI and NdeI sites at approximate positions 9,974 and

10,679, respectively. Note that mORF12-13 and mORF16-17 provide examples of combinations of a promoter and a polyadenylation site from two different T-DNA genes.

5 Example 4

This Example teaches the manipulation of four exemplary foreign structural genes preparatory for insertion into a pRI T_L-DNA TxCS. The genes are for the proteins phaseolin (a nutritionally important seed storage protein from Phaseolus vulgaris), P. vulgaris lectin (a nutritionally important protein found in seeds and other plant tissues which may be involved in symbiotic nitrogen fixation and making seeds unpalatable to herbivores), thaumatin (a protein which tastes sweet to primates, naturally found in Thaumatococcus daniellii), and crystal protein (a protein produced by Bacillus thuringiensis which is used commercially to control larval pests of a large number of lepidopteran insect species). The crystal protein structural gene used here, though lacking its 3' end, encodes a protein toxic to insect larvae. Phaseolin, lectin, and thaumatin are eukaryotic genes; crystal protein is prokaryotic. Phaseolin contains introns; lectin and crystal protein do not. The lectin gene itself contains no introns and could be obtained on a 5.7 kbp HindIII fragment from a genomic clone (L. M. Hoffman (1984) *J. Mol. Appl. Genet.* 2:447-453) which is part of a plasmid harbored by the deposited strain NRRL B-15821 (see also Example 6.4). However, in this Example the lectin structural gene is obtained from a cDNA clone (L. M. Hoffman et al. (1982) *Nucl. Acids Res.* 10:7819-7828), as is the thaumatin gene.

4.1 Subcloning structural genes into M13

The genes listed in Table 7, column 1 are carried by the plasmids listed in Table 7, column 2, which may be isolated from the deposited stains listed in Table 7, column 3 (e.g. the crystal protein structural gene is carried by p123/58-10 which is harbored within NRRL B-15612). DNA of a plasmid listed in Table 7, column 2 is digested to completion with the restriction enzyme(s) listed in the corresponding row of Table 6, column 4 and protruding ends are removed by incubation with the enzyme listed in Table 6, column 5 (e.g. for manipulation of the crystal protein structural gene, p123/58-10 DNA is digested with HindIII and the resulting

sticky-ends are removed by incubation with the Klenow fragment of E. coli DNA polymerase I). A DNA fragment whose size is listed in Table 7, column 6 (e.g. 6.6 kbp for the crystal protein) is isolated by elution from an agarose gel after electrophoretic separation. The resulting fragment is mixed with and ligated to dephosphorylated, blunt-ended, linearized m13(Nde), prepared as described in Example 3.1, and is transformed into E. coli. Viral DNAs and RFs are isolated from transformants and screened by restriction and hybridization analyses for the presence of inserts which are complementary to the sequence when in single-stranded viral form as present in the mRNA. The vector which infects the selected colony is designated as listed in Table 7, column 7 (e.g. mBtCP for the crystal protein).

4.2 Placement of NdeI sites flanking three structural genes

DNA of a vector listed in Table 8, column 1 is used to prepare a vector designated as listed in Table 8, column 2 by primer extension after hybridization to the oligonucleotides listed in Table 8, column 3 (e.g. for crystal protein, mBtCP is used to make mBtCP' by extending the primers 5'GGAGGTAACATATGGATAACAAATCCG3' and 5'GCGGCAGATTAACGTGTTCATATGCATTGAG3') as described in Examples 3.3 and 5. This has the effect of introducing NdeI sites at the translational start site and near the translational stop site; there are no BamHI or NdeI sites present within the structural gene which might otherwise be removed. In the case of the B. thuringiensis crystal protein gene, a translational stop codon (TAA) is additionally introduced. The structural genes listed in Table 7, column 1 may be isolated as a DNA fragment whose size is listed in Table 8, column 4 after digesting DNA of a vector listed in the corresponding line of Table 8, column 2 to completion with NdeI (e.g. the crystal protein structural gene is isolated from mBtCP' on a 2.8 kbp NdeI fragment).

4.3 Mutagenesis of thaumatin

Thaumatin cDNA-containing vectors have been disclosed by C. T. Verrips et al., Eur. Pat. applications 54,330 and 54,331, and L. Edens et al. (1982) Gene 18:1-12. Thaumatin is originally synthesized as preprothaumatin, the prefix "pre" representing the presence of a "signal peptide" having the function of causing the export of thaumatin

from the cytoplasm into the endoplasmic reticulum of the cell in which it is being synthesized, and the prefix "pro" representing that the protein is not in mature form. A thaumatin cDNA structural gene is present as the complement to thaumatin mRNA in M13-101-B (Eur. Pat. application 5 54,331). The viral form of this vector is used as a source of a thaumatin structural gene after site-specific mutagenesis directed by two of the following oligonucleotides: (a) 5'GGCATCATA~~CAT~~ATGGCCGCCACC3', (b) 5'CCTCACGCTCTCCGCGC~~AT~~ATGGCCACCTTCGAGATCGTCAACCGC3', (c) 5'CGAGTAAGAGGGATGAAGACGGAC~~AT~~TGAGGATAACGC3', or 10 (d) 5'GGGTCACTTTCTGCCCTACTGCC~~TAAC~~ATCAAGACGACTAAGAGG3'. When mutated by oligonucleotides (a) and (c), which bind to the 5'- and 3'-ends of the structural gene, respectively, a preprothaumatin sequence is extracted from the resultant vector by NdeI digestion. When mutated by oligonucleotides (b) and (d), which bind to the 5'- and 3'-ends, respectively, a 15 mature thaumatin sequence is similarly extracted. Use of the combinations of (a) with (d) and (b) with (c) yields fragments encoding what might be termed prethaumatin and prothaumatin, respectively. All of these sequences are obtained on fragments having a size of approximately 0.7 kbp having no internal NdeI or BamHI sites which may be isolated as usual by 20 gel electrophoresis.

4.4 Other possible manipulations

Phaseolin and lectin, as initially translated have signal peptides at their amino-termini, as is the case with thaumatin. If desired, these 25 signal peptides may be eliminated by placing the 5'-NdeI site between the codons forming the junction between the signal peptide and the mature protein. When under control of a T-DNA in a plant cell nucleus, such a structural gene will cause the synthesis of a phaseolin or lectin protein which is not exported from the cell's cytoplasm. Sequences useful for 30 designing oligonucleotides for manipulating for phaseolin and lectin structural genes are respectively reported by J. L. Slightom et al. (1983) Proc. Natl. Acad. Sci. USA 80:1897-1901, and Hoffman et al. (1982) supra.

Example 5

This Example describes techniques for the synthesis and use of synthetic oligonucleotides. Other useful references can be found in the list of works cited in the section introductory to these Examples.

5

5.1 Oligonucleotide synthesis

Techniques for chemical synthesis of DNA utilize a number of techniques well known to those skilled in the art of DNA synthesis. Modification of nucleosides is described by H. Schaller *et al.* (1953) J. Amer. Chem. Soc. 85:3621-3827, and H. Buchi and H. G. Khorana (1972) J. Mol. Biol. 72:251-288. Preparation of deoxynucleoside phosphoramidites is described by S. L. Beaucage and M. H. Caruthers (1981) Tetrahedron Lett. 22:1859-1862. Preparation of solid phase resin is described by S. P. Adams *et al.* (1983) J. Amer. Chem. Soc. 105:661-663. Hybridization procedures useful during the formation of double-stranded molecules are described by J. J. Rossi *et al.* (1982) J. Biol. Chem. 257:9226-9229.

10

15

20

25

30

35

5.2 Oligonucleotide-directed site-specific mutagenesis

General methods of directed mutagenesis have been reviewed by D. Shortle *et al.* (1981) Ann. Rev. Genet. 15:265-294. Of special utility in manipulation of genes is oligonucleotide-directed site-specific mutagenesis, reviewed recently by C. S. Craik (1985) Biotechniques 3:12-19; M. J. Zoller and M. Smith (1983) Meth. Enzymol. 100:468-500; M. Smith and S. Gillam (1981) in Genetic Engineering; Principles and Methods, Vol. 3, eds.: J. K. Setlow and A. Hollaender; and M. Smith (1982) Trends in Biochem. 7:440-442. This technique permits the change of one or more base pairs in a DNA sequence or the introduction of small insertions or deletions. Recent examples of oligonucleotide-directed mutagenesis include W. Kramer *et al.* (1984) Nucl. Acids Res. 12:9441-9456; Zoller and Smith (1983) *supra*; M. J. Zoller and M. Smith (1982) Nucleic Acids Res. 10:6487-6500; G. Dalbadie-McFarland *et al.* (1982) Proc. Natl. Acad. Sci. USA 79:6409-6413; G. F. M. Simons *et al.* (1982) Nucleic Acids Res. 10:821-832; and C. A. Hutchison III *et al.* (1978) J. Biol. Chem. 253:6551-6560. Oligonucleotide-directed mutation using double-stranded DNA vectors is also possible (R. B. Wallace *et al.* (1980) Science 209:1396-1400; G. P. Vlasuk *et al.* (1983) J. Biol. Chem. 258:7141-7148; E. D. Lewis *et al.*

(1983) Proc. Natl. Acad. Sci. USA 80:7065-7069; Y. Morinaga et al. (1984) Biotechnol. 2:636-639). See Example 3.1 for useful M13-based vectors.

Example 6

5 This Example teaches use of the pRi T_L -DNA TxCSs and the foreign structural genes manipulated in Example 3 and 4, respectively. Specific Examples of plant transformation vectors, plant transformation, and plant regeneration are given below in Examples 6.4-6.7.

10 6.1 Assembly of TxCS/structural gene combinations

A plasmid listed in Table 6, column 1 (e.g. mORF 11) is digested with NdeI and dephosphorylated with phosphatase, and the opened vector may be separated from the T-DNA structural gene found nested within the TxCS. A plasmid listed in Table 8, column 2 is digested with NdeI and the corresponding structural gene listed in Table 7, column 1 is isolated as a fragment whose size is listed in Table 8, column 4 by agarose gel electrophoresis followed by elution from the gel (e.g. crystal protein structural gene is isolated from mBtCP' on a 2.8 kbp NdeI fragment). Additionally, a thaumatin-encoding fragment may be isolated as described in Example 4.3.

20 Any desired combination of an opened TxCS vector and an isolated foreign structural gene may now be mixed with each other and ligated together. For example, crystal protein structural gene may be placed between an ORF 11 promoter and an ORF 11 polyadenylation site, thereby replacing the structural gene of ORF 11 with that of the crystal protein, by ligating the 2.8 kbp NdeI fragment of mBtCP' into NdeI-digested mORF 11 DNA. The ligation mixtures are individually transformed into E. coli and RFs are isolated from the resultant transformants and characterized by restriction analysis. A colony is chosen for each transformation which lacks the endogenous pRi T_L -DNA structural gene and has a single copy of the heterologous foreign structural gene inserted within the TxCS, the structural gene and the TxCS being in such orientation with respect to each other that the gene is expressible under control of the TxCS when within a plant cell.

6.2 Assembly of plant transformation vectors

A TxCS/heterologous foreign structural gene combination may be removed from the M13-based vector constructed in Example 6.1 by digestion with BamHI followed by agarose gel electrophoresis and elution. The size of the BamHI-fragment bearing the promoter/structural gene/polyadenylation site may be calculated by adding the size of the structural gene-bearing fragment, as listed in Table 8, column 4, to the sizes of the promoter and polyadenylation site-bearing segments, as listed in Table 6, column 6. For example, an ORF 11 TxCS/crystal protein structural gene combination, as exemplified herein, may be obtained on a 4.2 kbp BamHI fragment (2.8 kbp + 715 bp + 705 bp). A TxCS/gene combination may be inserted directly into a 5'GATC...3' sticky-ended site, which may be generated by BamHI, BclI, BglII, MboI, or Sau3AI. Alternatively, the combination may be inserted into any desired restriction site by conversion of sticky-ends into blunt-ends followed by blunt-end ligation or by use of appropriate oligonucleotide linkers.

An alternative to assembly of a pRI T_L-DNA TxCS/structural gene combination followed by insertion of that combination into a plant transformation vector is the insertion of a pRI TxCS into a plant transformation vector followed by insertion of the structural gene into the TxCS/transformation vector combination. It is advantageous that the plant transformation vector not contain NdeI sites if the particular manipulation strategy exemplified herein is to be used. Otherwise TxCS/vector combination may be linearized by partial NdeI digestion, as will be understood in the art.

6.3 Vector choice, transformation and plant regeneration

The plant transformation vector into which the TxCS/gene combination is to be inserted may be a TIP-based system such as a TIP plasmid, a shuttle vector for introduction of novel DNAs into TIP plasmids, or a sub-TIP plasmid, e.g. mini-Ti or micro-Ti. Alternatively, a vector based upon a DNA virus, minichromosome, transposon, and homologous or nonhomologous recombination into plant chromosomes may be utilized. Any mode of delivery into the plant cell which is to be initially transformed may be used which is appropriate to the particular plant transformation vector into which the TxCS/structural gene combination is inserted. These forms

of delivery include transfer from a Agrobacterium cell, fusion with vector-containing liposomes or bacterial spheroplasts, direct uptake of nucleic acid, encapsidation in viral coat protein followed by an infection-like process, or microinjection.

5 The initially transformed plant cells are propagated and used to produce plant tissue and whole plants by any means known to the art which is appropriate for the plant transformation vector and delivery mode being used. Methods appropriate for TIP-based transformation systems include those described by M.-D. Chilton *et al.* (1982) *Nature* 295:432-434, for carrots, K. A. Barton *et al.* (1983) *Cell* 32:1033-1043, for tobacco. 10 Selection of transformed cells may be done with the drugs and selectable markers as described in the Background. The exact drug, concentration, plant tissue, plant species and cultivar must be carefully matched and chosen for ability to regenerate and efficient selection. Screening of 15 transformed tissues for tissues expressing the foreign structural gene may be done using immunoassays known to the art. Southern, northern, and dot blots, all methods well known to those skilled in the art of molecular biology, may be used to detect incorporated or expressed nucleic acids. Screening for opine production is also often useful.

6.4 Preparation of a disarmed T-DNA vector

E. coli C600 (pRK-203-Kan-103-Lec), which is on deposit as NRRL B-15821, is a pRK290 derivative containing T-DNA sequences of pTi15955 from between EcoRI sites at positions 4,494 and 12,823, as defined by R. F. Barker et al. (1983) Plant Mol. Biol. 2:335-350, except for a deletion of sequences between position 5,512 HindIII site and position 9,062 BamHI site. Inserted into the deletion, i.e. substituting for the deleted T-DNA, is a Tn5-derived kanamycin resistance (kan) gene and a Phaseolus vulgaris seed lectin gene (see Example 4, Hoffman (1984) supra). The lectin gene may be deleted from pRK-203-Kan-103-Lec by digestion with HindIII followed by religation; the resultant vector is designated pRK-203-Kan-103. BamHI-digested, dephosphorylated pRK-203-Kan-103 is mixed with and ligated to a BamHI fragment bearing the pRi T_L-DNA T_{CS}/heterologous foreign structural gene combination assembled in Example 6.2; the resultant vector is designated pRK-203-Ri-Kan-103. pRK-203-Ri-Kan-103 is introduced in A. tumefaciens ATCC15955 using methods

well known in the art, and a double-homologous recombinant, designated RS-Ri-Kan, is identified. RS-Ri-Kan does not harbor pRK-203-Ri-Kan-103, but contains a mutated pTi15955 having a T-DNA substitution between the positions 5,512 HindIII site and 9,062 and BamHI site of a TxCS/structural gene combination and a kan gene for pTi T-DNA. This substitution deletes some tmr and tms sequences, thereby disarming the T-DNA. RS-Ri-Kan T-DNA transforms inoculated plant tissue without conferring the phenotype of hormone-independent growth. Tobacco tissues transformed by RS-Ri-Kan may be regenerated into normal plants using protocols well known in the art for regeneration of untransformed tissue.

6.5 Construction of a micro-Ti plasmid

p102, a pBR322 clone of the pTi15955 T-DNA fragment between HindIII sites at positions 602 and 3,390 (as defined by R. F. Barker *et al.*, *supra*) carries the left border of T_L and promoter sequences associated with ORF 1. p233 is a pBR322 clone of the pTi15955 T-DNA BamHI/EcoRI fragment spanning positions 9,062 and 16,202. The T-DNA of p233 includes a SmaI/BclI fragment spanning positions 11,207 and 14,711, having ocs, a 3'-deleted tml, and the right border of T_L . p233 was linearized with SmaI, mixed with and ligated to a commercially available blunt-end BglII linker, trimmed with BglII, religated to itself, and transformed into *E. coli* GM33 (a dam⁻ host that does not methylate DNA in a manner incompatible with the action of BclI, M. G. Marinus and N. R. Morris (1974) *J. Mol. Biol.* 85:309-322). A colony was identified which harbored a plasmid, designated p233G, having a BglII site in the location formerly occupied by the position 11,207 SmaI site. p233G DNA was digested with BglII and BclI and a 3.5 kbp fragment was isolated by agarose gel electrophoresis followed by elution. The 3.5 kbp BglII/BclI fragment was mixed with and ligated to BclII-digested, phosphatase-treated p102 DNA. The ligation mixture was transformed into *E. coli* K802 (W. B. Wood (1966) *J. Mol. Biol.* 16:118). Plasmid DNAs from ampicillin-resistant transformants were characterized by restriction analysis and a colony was identified, designated pAK-4, having the BglII/BclI fragment of p233G inserted into the BclII site of p102 and oriented so that the ocs gene was located between the left and right T_L borders. One BglII site, also between the borders, was regenerated, and a BglII/BclI suture, not susceptible to the

action of either enzyme, was generate to the right of the right border. pAK-4 may be represented as follows:

....pBR322...HindIII...left border...BglII...ocs...right border...
(BglII/BclI)...HindIII...pBR322....

5

The T-DNA of pAK-4 may be removed on a 6 kbp HindIII fragment. HindIII-digested pAK-4 DNA was mixed with and ligated to HindIII-linearized, phosphatase-treated pSUP106 DNA. pSUP106, a 10 kbp wide host-range plasmid capable of maintenance in both E. coli and Agrobacterium (R. Simon *et al.* (1983) in Molecular Genetics of the Bacteria-Plant Interaction, ed.: A. Pühler, pp. 98-106), is harbored by E. coli CSH52 (pSUP106) which is on deposit as NRRL B-15486. The reaction mixture was transformed into K802 and plasmid DNAs from chloramphenicol-resistant transformants were characterized by restriction analysis. A colony was identified harboring a plasmid, designated pAN6, having the Agrobacterium DNA of pAK-4 inserted into the HindIII site of pSUP106 oriented so that BglII/BclI suture was proximal to the pSUP106 EcoRI site. pAN6 is a micro-Ti plasmid having within its two T-DNA borders a functional ocs gene and a BglII site that is unique to the plasmid. The BglII site is flanked by an incomplete tml gene and the pTi ORF 1 promoter, both of which are transcribed towards the BglII site.

BamHI-digested, dephosphorylated pAN6 is mixed with and ligated to a BamHI fragment bearing the pRi T_L-DNA TxCS/heterologous foreign structural gene combination assembled in Example 6.2; the resultant vector is designated pAN6-Ri. pAN6-Ri may be introduced into an Agrobacterium strain having a helper plasmid, e.g. LBA4404 (G. Ooms *et al.* (1981) Gene 14:33-50), using methods well known in the art.

6.6 Inoculation of tobacco stems

Stems of sterile Nicotiana tabacum var. Xanthi are cut into segments approximately 1 cm long. These segments are placed basal end up in Petri dishes containing Murashige and Skoog medium (MS medium: 1.65 g/l NH_4NO_3 , 1.9 g/l KNO_3 , 440 mg/l $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 370 mg/l $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 170 mg/l KH_2PO_4 , 0.83 mg/l KI, 6.2 mg/l H_3BO_3 , 22.3 mg/l $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$, 8.6 mg/l $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.25 mg/l $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.025 mg/l $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.025 mg/l $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 37.23 mg/l Na_2EDTA , 27.85 mg/l $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g/l inositol, 50 mg/l

nicotinic acid, 50 mg/l pyroxidine-HCl, 50 mg/l thiamine-HCl, 30 g/l sucrose, and 8 g/l agar, pH 5.8) without hormonal supplement, a medium well known in the art. The basal (upper) ends are then inoculated with Agrobacterium cells by puncturing the cut surface of the stem with a 5 syringe needle. After two weeks of incubation at 28°C with 16 hr light and 8 hr dark, calli develop at the upper surface of all stem segments. The callus regions are then transferred to MS medium containing 2.0 mg/l NAA (1-naphthalene acetic acid), 0.3 mg/l kinetin and 0.5 mg/ml carbini- 10 cillin. After two weeks on this medium, the tissues are free of bacteria and can be assayed for the presence of opines, a methodology well known in 15 the art.

Once free of inciting bacteria, the transformed plant tissues are grown on MS medium with hormones at 25°C with 16 hr light and 8 hr dark. These tissues are cloned using a suspension method described by A. N. 15 Binns and F. Meins (1979) *Planta* 145:365-369. Briefly, tissues are placed in liquid MS medium supplemented with 2.0 mg/l NAA and 0.1 mg/l kinetin, and shaken at 135 rpm at 28°C for 10-14 days. The resultant suspensions are filtered successively through 0.543 and 0.213 mm mesh sieves, concentrated, and plated at a final density of 8×10^3 cells/ml in MS medium 20 supplemented with 2.0 mg/l NAA and 0.3 mg/l kinetin. After these grow to approximately 100 mg, colonies are split into two pieces. One piece is placed on complete MS medium and the other is screened for the presence of 25 opines. Approximately 0-50% of the colonies are found to be opine-positive, depending on the particular parental uncloned callus piece from which the colonies were descended. Uncloned pieces having higher concentrations of opine tended to yield a higher percentage of opine-positive clones.

6.7 Regeneration of recombinant plants

30 Tissues from various opine-positive clones are transferred onto MS medium supplemented with 0.3 mg/l kinetin and cultured at 28°C with 16 hr light and 8 hr dark. Shoots initiated are subsequently rooted by placing them in MS medium without hormones. Rooted plantlets are transferred to soil and placed at high humidity in a greenhouse. After 7-10 days, the 35 plants are then grown with normal greenhouse conditions. Regenerated

plants derived from opine-positive clones contain opines. The presence of opines indicates thereby that these normal looking plants are transformed by T-DNA.

Table 1 Restriction Enzyme Sites in pR1 T₁-DNA Region

		No. Sites	Locations			
5	Bst E II	1	3 993			
	Sna I	1	6 459			
	Apa I	2	3 390	17 851		
10	Mst II	2	4 806	15 021		
	Sma I	2	3 075	9 863		
	Xba I	2	676	4 999		
	Kpn I	3	3 364	14 133	19 918	
15	Mlu I	3	17 606	20 793	20 856	
	Nco I	3	2 262	10 133	21 021	
	Sst II	3	3 431	14 691	17 037	
	Xba I	3	9 242	11 003	20 700	
20	Bam HI	4	1 343	11 198	11 278	12 816
	Hpa I	4	8 375	12 459	13 700	18 818
	Nde I	4	3 519	3 861	4 822	10 308
25	Nru I	4	5 281	10 968	11 617	18 901
	Sal I	4	4 515	6 047	12 655	15 821
	Ava III	5	13 684	14 382	15 480	16 415
	BssH II	5	5 727	6 847	19 761	20 260
30	BstX I	5	2 269	4 226	9 912	16 016
	Cla I	5	35	753	11 421	12 598
	Nar I	5	465	4 114	11 356	16 441
	Nsi I	5	13 688	14 386	15 484	16 419
	Sca I	5	1 794	4 546	10 166	11 500
35	Tth III I	5	3 413	3 816	8 217	8 769
	Xma III	5	5 814	7 970	8 502	10 613
	Aat II	6	974	5 615	6 054	7 521
	Asu II	6	4 792	10 026	12 954	16 897
40	Hind III	6	5 602	6 361	9 814	11 587
	Mst I	6	4 004	8 091	11 427	16 088
	Pst I	6	2 244	4 892	7 003	10 486
	Xba II	6	230	2 659	4 480	5 694
	Bcl I	7	992	1 364	6 710	10 564
			19 827			
	Bgl II	7	4 197	5 525	7 879	11 239
			15 760			
	EcoR I	7	7 585	9 077	13 445	15 358
			18 911			
					17 059	18 766

Table 1 continued

		No. Sites	Locations					
5	Enzyme							
	Acc I	8	333	4 516	6 048	6 460	9 514	12 656
10			15 822	19 089				
	Bal I	8	497	3 568	5 488	9 233	9 339	9 916
			12 001	17 544				
	Sph I	8	582	11 476	15 013	15 057	15 486	17 175
			19 027	20 404				
15	Xmn I	8	1 759	2 725	4 498	4 546	10 103	12 206
			17 338	17 917				
	EcoR V	9	5 134	6 738	7 775	10 098	10 626	13 173
			14 048	16 080	17 491			
	Sst I	9	1 967	4 152	10 879	11 068	12 395	14 105
			17 016	19 214	19 866			
20	Stu I	9	5 590	6 696	7 512	11 442	12 066	15 967
			16 656	20 186	20 467			
	Bgl I	10	1 571	3 125	5 872	5 956	6 832	9 775
			10 912	14 290	16 606	21 065		
25	Ava I	11	3 073	3 765	5 268	7 012	9 242	9 861
			10 573	10 629	11 003	14 402	20 700	
	Aha III	12	2 486	11 334	12 233	13 427	13 580	13 666
			15 577	15 599	16 168	18 135	18 573	20 070
	Nae I	13	316	446	1 664	3 931	3 962	5 733
30			7 616	9 771	15 000	16 622	18 474	20 380
			20 652					
	Pvu II	13	250	1 235	1 859	2 395	2 752	7 888
			8 451	12 042	13 715	15 590	15 620	16 056
			18 688					
35								
	Ban II	19	Hph I	37	Hpa II	72		
	HgiA I	19	Rsa I	38	Cfo I	80		
	Ban I	20	HinF I	41	HinP I	80		
10	Hinc II	21	Hga I	42	Ala I	87		
	Xho II	22	Fok I	48	Sau 3a	87		
	Hae II	23	Dde I	55	Hae III	99		
	Nci I	23	Mbo II	63	Taq I	113		
	Aha II	24	Sau 96	66	Fnu 4A	132		
45	Ava II	26	Fnu II	68	Mnl I	171		
	BstN I	35	Bbv I	69				

Table 2. Open-Reading Frames in pRi T₁-DNA

ORF	Sequence location		Coding sequence base pairs	Amino acids	Calculated molecular weight (daltons) of ORF-encoded protein
	After first ATG in-frame	Before terminator			
1.	22262	937	GGC ATG G	1326	442
2.	3458	2649	GAT ATG T	810	270
3.	3726	4799	ATC ATG C	1074	358
4.	4100	4041	GGG ATG C	360	120
5.	4918	4607	GGG ATG C	312	104
6.	5143	6216	CGT ATG C	1074	358
7.	5643	5071	GCC ATG G	573	191
8.	6609	8888	GTG ATG G	2280	760
9.	6830	6576	GCC ATG A	255	85
10.	9748	10044	AGA ATG G	297	99
11.	11282	10509	ACA ATG G	774	258
12.	12166	13002	AAC ATG G	537	179
13.	13723	14319	TGA ATG G	597	199
14.	15659	16210	AGC ATG G	552	184
15.	17545	16517	CAG ATG G	1029	343
16.	18189	17737	AAA ATG T	153	151
17.	18743	18177	GAG ATG A	567	189
18.	19390	19031	AAC ATG G	360	120

Coordinates represent the λ of the AUG initiation codon or the last nucleotide before the termination codon.

Table 3. Eukaryotic Transcription Controlling Sequences

ORF	Sequence and position of promoter elements, positions from first ATG	Sequence and position of polyadenylation sites, positions from terminator	Distance (bp) from first Met to best polyadenylation site
1.	(CCMT) -211 ; CAAAT	(TATA) -143 ; ATAA ; TATA -92 ; ATAA ; TATA -65	+38 ATATA ; ATATA +100
2.	-81 CCAAT	-60 ATAT	+116 ATAAA
3.	-102 CAACT	-80 TATA	+137 ATGAA
4.	-107 CCAAA	-82 -16 ATAAA ; ATAA	+380 ATAAA
5.	-131 CCAAAAT	-68 ATAA	+119 GATAAA
6.	-146 CAAAAT	-98 ATATA	+97 ATATA ; ATATA +260 ATAAA ; ATAAA +294
7.	---	---	---
8.	-133 CCTACCA ; CAAAGT	-92 TAAAT ; TATA -72	+96 ATAAA ; ATAAA +236
9.	-76 CAATT	-59 TATA	+187 ATAAA
10.	-221 CATAT ; CAAAT ; CAAAT	-144 TATA ; TATA -25	+75 ATAGS ; ATATA +114
11.	-116 CCAAA	-54 TATA	+350 ATAAA

Table 3 continued

ORF	Sequence and position of promoter elements, positions from first ATG		Sequence and position of polyadenylation sites, distance from terminator		Distance (bp) from first Met to best polyadenylation site	
12.	-81 CCATT		-56 TATATA	+83 ATTAAA ; ATAAA	+141 620	
13.	-155 CAATT		-87 ATATA ; TATA	+111 ATATA ; ATATA	+262 708	
14.	-174 CCATT ; CAAA ; CAGG	-95	-140 ATAA ; TATAA ; ATA	-72 ATAAA ; ATATA ; ATATA	-50 +60 +128 +231	612
15.	-91 CCAAA		-65 TATATA	+149 ATATA	1178	
16.	-193 CAAA		-126 TATA	+87 ATTAAA ; TATA	+120 545	
17.	-69 CAATC	-50	-60 ATAT ; ATATA	-37 +92 TATATA ; ATATA	+164 670	
18.	---	---	---	---	360	

Element positions are negative or positive when respectively 5' or 3' to an ORF.

Table 4 Construction of pR1 T_L-DNA ORF-carrying vectors based on M13

1a	2b	3b	4b	5c
to use ORF 1: e36	<u>Sma</u> I	3.8 kbp	ORF 1	mR1
to use ORFs 2-9: pLJ40 or <u>Bam</u> H 8a	<u>Nco</u> I	7.9 kbp	ORFs 2-9	mR2
to use ORF 10: pLJ40 or <u>Bam</u> H 8a	<u>Bgl</u> II and <u>Bam</u> HI	3.3 kbp	ORF 10	mR3
to use ORFs 11-13: pLJ40 or e16	<u>Sma</u> I and <u>Mst</u> II	5.2 kbp	ORFs 11-13	mR4
to use ORFs 14-17: pLJ40	<u>Kpn</u> I	5.8 kbp	ORFs 14-17	mR5

a plasmids listed in column 1 are used as sources of pR1 T_L-DNA sequences.

bAfter restriction enzymes listed in column 2 are used to cut the plasmids listed in column 1, DNA fragments having sizes listed in column 3 are isolated which carry the ORF(s) listed in column 4.

cThe resultant M13-based vectors designated in column 5 carry the ORFs designated in column 4.

Table 5 Removal of endogenous NdeI and BamHI sites from pRI T_L-DNA from pRI T_L-DNA

	1 ^a	2 ^a	3 ^b	4 ^b	5 ^b
to use ORF 1:	mR1'	mR1	5' CTGATGACTACGG <u>AGCCTGGACAGC</u>	3'	<u>BamHI</u> 1,343
to use ORFs 2-9:	mR2'	mR2	GCACGCCACAGAT <u>GTAGATGAGATGAG</u> GGTCCTAGGGC <u>GTATGTAATGTTTGC</u>	<u>NdeI</u> <u>BamHI</u>	3,518 3,860
to use ORF 10:	mR3'	mR3	GATTAGATAGT <u>CAAGATGAGCATGTGTC</u>	<u>NdeI</u>	4,821
to use ORFs 11-13:	mR4'	mR4	GATTAGATAGT <u>CAAGATGAGCATGTGTC</u> GC AAATCGG <u>GGCCCTCGAAATAGG</u> GC AAATGGGAGC <u>CCATTGTGATGTG</u> CGGTTACGGG <u>GGCTATGGGGAGCCGC</u>	<u>NdeI</u> <u>BamHI</u>	10,305 11,198

^aModified vectors having designations listed in column 1 are made from the vectors listed in column 2.

^bOligonucleotides listed in column 3 are used as primers in site-specific mutagenesis of vectors listed in column 2, thereby removing restriction sites specific to the enzymes listed in column 4 which are found in the pRI T_L-DNA sequence near the positions listed in column 5. Underlined letters in column 3 indicate positions of introduced mutations.

Table 6 Placement of NdeI and DamII sites flanking pRI T_L-DNA TxCS elements

	1 ^a	2 ^a	3 ^b	4 ^b	5 ^c	6 ^c
to use ORF 1:	mORF1	mR1'	5'	CTAGAGACCCGGATGGATCGTATAGTCAGCACCG GGCTCTGGTGCCTATATGACGTCGAGATGAGG CCTATCTTACCCATATGCAATGGGTTTGC GTTAACAGGAGGGAT <u>CCGGAA</u> ACAGTGC	3'	692 A 275 bp 967 A 275 bp 2,265 P 346 bp 2,611
to use ORF 2:	mORF2	mR2'		GATAGGGGGATCCCGTTATTA <u>CTGTTCTCC</u> CGGCCATT <u>CTGTTCTCC</u> ATGAGGAGAG GAGACATAT <u>GGATCC</u> AGAGGTGGGG GTTGCAGGCA <u>ATGGATCC</u> CACAGGGT	2,324 A 255 bp 2,579 P 411 bp 3,474 P 411 bp 3,885	
to use ORF 3:	mORF3	mR2'		CCCTTGAGGGATCC <u>CTGGTCAA</u> ATTGG CCTCGCTAT <u>CTATG</u> CCGCCACGAGGG GGTCGTCA <u>AGGCAT</u> GTCTGGAC CCGAGTGGGAT <u>CCCTTG</u> GGCTTG	3,281 P 442 bp 3,723 A 482 bp 4,821 A 285 bp 5,106	
to use ORF 4:	mORF4	mR2'		CGCAGACATATGGAT <u>CCGGT</u> GA <u>CTTGG</u> GCCACCCGACAT <u>ATGG</u> CC <u>AGTCGGATG</u> CCGCCACCGAT <u>TTAACCA</u> AT <u>GGGG</u> GG <u>TCGAGG</u> C <u>CGAGGAGATGGCC</u> TC <u>GGAT</u> <u>CC</u> TC <u>GGCC</u> CT <u>CCCC</u>	3,531 A 494 bp 4,025 A 494 bp 4,405 P 352 bp 4,757	
to use ORF 5:	mORF5	mR2'		CAGTGACCCACGGAT <u>CCGGAA</u> AGTCATTGG CTGTCCTCCATAT <u>GG</u> AT <u>GG</u> AT <u>GG</u> CGTTCCGT <u>GG</u> AC <u>GGCA</u> AT <u>GG</u> AT <u>GG</u> AT <u>GG</u> GG <u>GGCA</u> AC <u>GGG</u> AT <u>GG</u> AT <u>GG</u> AT <u>GG</u>	4,277 A 412 bp 4,689 A 412 bp 4,923 P 379 bp 5,302	
to use ORF 6:	mORF6	mR2'		CGACGGCTGGAT <u>CCCTGT</u> TC <u>CC</u> CGTGC <u>ACGGAT</u> AT <u>GG</u> AT <u>GG</u> CA <u>ATCCAC</u> AT <u>GG</u> AT <u>GG</u> AC <u>GG</u> AT <u>GG</u> GCT <u>CGGTGATGG</u> AT <u>GG</u> AT <u>GG</u>	4,675 P 463 bp 5,138 P 463 bp 6,221 A 392 bp 6,613 A 392 bp	

Table 6 continued

to use ORF 8:	mORF8	mR2'	CGTTCATCTGGAGAGGATCCAGACG CCTGAGTGGCTGTATATGGATCTTCATATTCAGCG 6,156 P 448 bp
			6,604 P 316 bp
			6,917 A 316 bp
			9,233
to use ORF 9:	mORF9	mR2'	CGTTCATCTGGAGAGGATCCAGACG GGATCTTCCATATGGCCGCCACG GCACCTGCCTCATATGGCCCCGAAATGG CCATGATGAACGTGATCCAGCAGACTGC 6,156 P 465 bp
			6,621 A 465 bp
			6,833 P 406 bp
			7,239
to use ORF 10:	mORF10	mR3'	CTTTTAAATTCTTGGATCCACTGGCCATTAAATTC CCTATTCTATGTTCAATGGATTAGCC CGGGATTAAATACGCAATGGCTGGCG CAGCAGGAGGATCCACAGAGAA 9,332 P 411 bp
			9,743 A 295 bp
			10,054 A 295 bp
			10,349
to use ORF 11:	mORF11	mR4'	GCTGGGAAAGGATCCCTTTGTCGCC GGCAGGCTAACATCATATGGGGGG GGGATCCAAATGTAATGAGTTGG GCCTTAAAGGAAATGGTGAATCCATGTACGTGC 9,974 A 705 bp
			10,679 P 715 bp
			11,286 P 715 bp
			12,001
to use ORF 12:	mORF12	mR4'	GCCTTAAAGGAAATGGTGAATCCATGTACGC CCTACTTGTAAATGGCTGAGACGACC GACATTGACCTATGGCAAGGCC CTTCTGAAAGAGGATCCGACATGTTTC 12,001 P 462 bp
			12,463 A 421 bp
			12,930 A 421 bp
			13,351
to use ORF 13:	mORF13	mR4'	CTCTGAAAGGAAATGGTGAATCCGACATGTTTC CCTTAAAGTGGCAAGGAAATGGCTCGTTATGGAGTGG GCTTATGACGCCATGGCAATCGGC CGTTTATGGGAGGATCCGGGGCCG 13,718 P 367 bp
			14,290 A 397 bp
			14,607
to use ORF 14:	mORF14	mR5	GCCTTAAAGGAAATGGGATCCACTAGG GGGAGAGTAAAGCATATGGCAAGATGGTTGG GTGATATGTTTACATGGAGGAG CCAGGCGAGTTGGATCTAACCTAAATTTC 15,242 P 414 bp
			15,656 A 321 bp
			16,170 A 321 bp
			16,491

Table 6 continued

to use ORF 15:	mORF15	mR5	CGTTTGAAACGGATCCGGGGTATG CGGCAACGCTCAATGCTTGTAGCC GTTGTGTTGGCCATATGCCCTGTAGG <u>CGCAAGGGTAGTCGA</u> TCCCAAGGCCGC	16,252 } A 338 bp 16,590 } 17,549 } P 339 bp 17,888 }
to use ORF 16:	mORF16	mR5	GCTTTGACATGGATCCGGCTTTCTGC <u>CCGGTCAGA</u> ACATGGATGAGTTGTC <u>CTGTGATCTCC</u> TTATGAAAGACGGGCTGTTG <u>CATTGATCTGG</u> CTGGAT <u>CC</u> CTGCATAGAG	17,416 } A 325 bp 17,741 } 18,748 } P 254 bp 19,002 }
to use ORF 17:	mORF17	mR5	<u>GGTAGTCGG</u> ATCCGGGGCGAAC <u>GGTACATTTTCACATATGCTGT</u> CCCC <u>GTCTGTGATCTCC</u> TTATGAAAGACGGGCTGTTG <u>CGGATTAGAATG</u> AT <u>CC</u> TTTCACACC	17,887 } A 314 bp 18,201 } 18,748 } P 424 bp 19,172 }
to use the ORF 12 promoter with the ORF 13 polyadenylation site:	mORF12-13	mR4'	<u>GCCTTAA</u> AGGGATGGGATCCATGTACGC <u>CCTACTTTGTT</u> ACATGGCTGAAAGACGACC <u>GCTTATGACGG</u> CAATGGCAATCGGC <u>CGTTTATGGAGG</u> ATCCGGGGCCG	12,001 } P 462 bp 12,463 } 14,290 } A 397 bp 14,687 }
to use the ORF 16 promoter with the ORF 17 polyadenylation site:	mORF16-17	mR5	GCTTTGACATGGATCCGGCTTTCTGC <u>CCGGTCAGA</u> ACATGGATGAGTTGTC <u>GTCTGTGATCTCC</u> TTATGAAAGACGGGCTGTTG <u>CGGATTAGAATG</u> AT <u>CC</u> TTTCACACC	17,416 } A 325 bp 17,741 } 18,748 } P 424 bp 19,172 }

^aModified vectors having designations listed in column 1 are made from the vectors listed in column 2.

^bOligonucleotides listed in column 3 are used as primers in site-specific mutagenesis of vectors listed in column 2, thereby placing NdeI and BamII sites flanking each ORF at a position equivalent to the pRI T₁-DNA position listed in column 4. The first and fourth oligonucleotide of each quartet introduce BamII sites, while the second and third introduce NdeI sites. Underlined and lower case letters in column 3 indicate positions of introduced mutations.

^cPairs of oligonucleotides marked in column 5 with "p"s define promoter-bearing segments as exemplified herein, while pairs marked with "A" define polyadenylation site-bearing segments, the segments having approximate sizes indicated in column 6.

Table 7 Construction of vectors carrying structural genes

to use	1 ^a	2 ^a	3 ^a	4 ^b	5 ^b	6 ^c	7 ^c
phaseolin:	p8.8	NRRL B-15393	<u>Bam</u> H I and <u>Bgl</u> II	Klenow fragment of <u>E. coli</u> DNA polymerase I	3.8 kbp	m ^{ph} as	
to use lectin:	ppVIL134	ATCC 39181	<u>Pst</u> I	bacteriophage T4 DNA polymerase	0.95 kbp	m ^{lec}	
to use crystal protein:	p123/58-10	NRRL B-15612	<u>Hind</u> III	Klenow fragment of <u>E. coli</u> DNA polymerase I	6.6 kbp	m ^{8tCP}	

^aStructural genes encoding the proteins listed in column 1 are carried by plasmids listed in column 2 which are harbored by the deposited strains listed in column 3.

^bDNA_s of plasmids listed in column 2 are digested with the restriction endonuclease(s) listed in column 4 and incubated with the enzymes listed in column 5 to convert sticky-ends to blunt-ends.

CDNA fragments of the sizes listed in column 6 are isolated and combined with an M13-based vector describe Example 3.1 to form the vectors listed in column 7.

Table 8 Placement of NdeI sites flanking structural genes

	1a	2a	3b	4c
to use phaseolin:	mphas	mphas'	5'	CTACTCTACATATGAGGAGGG GTAGGGTGTAGAGCTCA <u>T</u> GGAGAGCAGTGG 3'
to use lectin:	mLec	mLec'		GTAGGGTGTAGAGCTCA <u>T</u> GGAGAGCAGTGG 3'
to use crystal protein:	mBtCP	mBtCP'		GGAGGGTACATATGATCATGGCTCCCTCC CCTGCTAAT <u>M</u> GTC <u>T</u> ATGTCACAC GGGGCAGAT <u>M</u> G <u>T</u> TC <u>T</u> ATGCA <u>T</u> CG 2.8 kbp

⁵
a The vectors listed in column 1 are used to make the modified vectors listed in column 2.

⁶
b Oligonucleotides listed in column 3 are used as primers in site-specific mutagenesis of vectors listed in column 1, thereby placing NdeI sites flanking each structural gene. Underlined letters indicate positions of introduced mutations.

c Structural genes carried by vectors listed in column 2 may be isolated on NdeI fragments having sizes listed in column 4.

CLAIMS

We claim:

- 5 1. A method of genetically modifying a plant cell comprising the step of transforming the cell to contain a pRi T-DNA promoter and a heterologous foreign structural gene, the promoter and the structural gene being in such position and orientation with respect to one another that the structural gene is expressible in a plant cell under control
10 of the promoter.
2. A method according to claim 1, wherein the pRi T-DNA is hybridizable to pRiHRI T_L-DNA
- 15 3. A method according to claim 2, wherein the T-DNA promoter is from a gene selected from a group consisting of genes for ORFs 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17.
4. A method according to claim 3, wherein the T-DNA gene is selected from a group consisting of genes for ORFs 1, 2, 3, 6, 8, 11, 12, 13, 14, 15, and 16.
- 20 5. A method according to claim 4, wherein the T-DNA gene is selected from a group consisting of genes for ORFs 8, 11, 12, 13, and 15.
6. A method according to claim 2, wherein the T-DNA gene is from pRiHRI T-DNA, pRiA4 T-DNA, or a T-DNA essentially identical thereto.
- 25 7. A method according to claim 1, wherein the cell is additionally transformed to contain a pRi T_L-DNA transcript terminator, the promoter, the structural gene, and the transcript terminator being in such position and orientation with respect to one another that transcriptional termination of the structural gene in a plant cell is under control of the transcript terminator.
8. A method according to claim 1, wherein the promoter or the structural
30 gene comprises an insertion, deletion, or substitution of one or more nucleotide pairs.
9. A method according to claim 1, wherein the structural gene changes a phenotype of a plant or plant cell when expressed therein.
10. A method according to claim 9, wherein the structural gene encodes an
35 insecticidal toxin identical to or derived from the crystal protein of Bacillus thuringiensis.

11. A method according to claim 9, wherein the structural gene is hybridizable to a phaseolin gene.
12. A method according to claim 9, wherein the structural gene encodes thaumatin or a precursor of thaumatin.
- 5 13. A method according to claim 9, wherein the structural gene encodes a legume lectin.
14. A method according to claim 1, comprising the step of integrating the promoter/structural gene combination into a plant chromosome, whereby the combination is flanked by plant DNA.
- 10 15. A plant, plant cell, or plant tissue, or plant seed derived or descended from a genetically modified plant cell produced by the method of claim 14.
16. A plant, plant cell, or plant tissue, or plant seed derived or descended from a genetically modified plant cell produced by the method of claim 1.
- 15 17. A DNA molecule comprising a pRi T-DNA promoter and a heterologous foreign structural gene, the promoter and the structural gene being in such position and orientation with respect to one another that the structural gene is expressible in a plant cell under control of the promoter.
- 20 18. A DNA according to claim 17, wherein the pRi T-DNA is hybridizable to pRiHRI T_L-DNA.
19. A DNA according to claim 18, wherein the T-DNA promoter is from a gene selected from a group consisting of genes for ORFs 1, 2, 3, 4, 25 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17.
- 20 20. A DNA according to claim 19, wherein the T-DNA gene is selected from a group consisting of genes for ORFs 1, 2, 3, 6, 8, 11, 12, 13, 14, 15, and 16.
21. A DNA according to claim 20, wherein the T-DNA gene is selected from a group consisting of genes for ORFs 8, 11, 12, 13, and 15.
- 30 22. A DNA according to claim 18, herein the T-DNA gene is from pRiHRI T-DNA, pRiA4 T-DNA, or a T-DNA essentially identical thereto.
23. A DNA molecule according to claim 17, further comprising a pRi T_L-DNA transcript terminator, the promoter, the structural gene and the transcript terminator being in such position and orientation with respect to one another that transcriptional termination of the

structural gene in a plant cell is under control of the transcript terminator.

24. A DNA according to claim 17, wherein the promoter or the structural gene comprises an insertion, deletion, or substitution of one or more nucleotides.
25. A DNA according to claim 17, wherein the structural gene changes a phenotype of a plant or plant cell when expressed therein.
26. A DNA according to claim 25, wherein the structural gene encodes an insecticidal toxin identical to or derived from the crystal protein of Bacillus thuringiensis.
27. A DNA according to claim 25, wherein the structural gene is hybridizable to a phaseolin gene.
28. A DNA according to claim 25, wherein the structural gene encodes thaumatin or a precursor of thaumatin.
29. A DNA according to claim 25, wherein the structural gene encodes a legume lectin.
30. A DNA according to claim 17, wherein the DNA is contained within a bacterium.
31. A DNA according to claim 30, wherein the bacterium is E. coli or is of the genus Agrobacterium.
32. A DNA according to claim 17, wherein the DNA is within a plant cell.
33. A DNA according to claim 32, wherein the plant cell is within a plant, a plant tissue, or a plant seed.
34. A DNA according to claim 17, wherein the promoter/structural gene combination is flanked by plant DNA.
35. A DNA according to claim 34, wherein the DNA is within a plant cell, a plant tissue, a plant, or a plant seed.
36. A DNA according to claim 17, wherein the DNA is within a plant cell, a plant tissue, a plant, or a plant seed.
37. A DNA molecule comprising a heterologous foreign structural gene and a pRI T_L-DNA transcript terminator, the structural gene and the transcript terminator being in such position and orientation with respect to one another that transcriptional termination of the structural gene in a plant cell is under control of the transcript terminator.

38. A DNA according to claim 37, wherein the transcript terminator is derived from a gene selected from a group consisting of genes for ORFs 1, 2, 3, 6, 8, 11, 12, 13, 14, 15, and 16.

5 39. A DNA according to claim 38, wherein the T-DNA gene is selected from a group consisting of genes for ORFs 8, 11, 12, 13, and 15.

40. A DNA according to claim 37, wherein the DNA is within a plant cell, a plant tissue, a plant, or a plant seed.

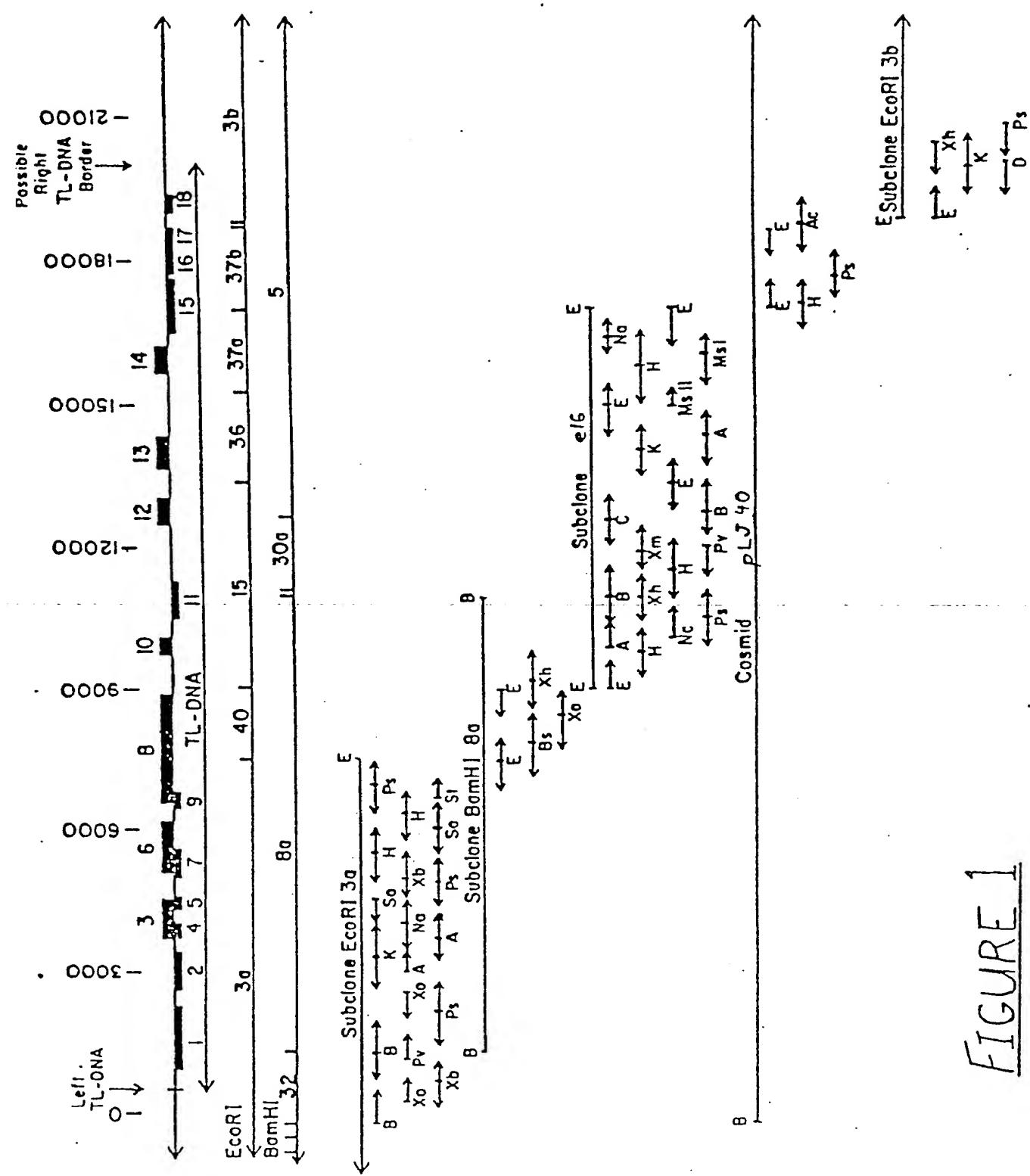


FIGURE 1

GGCCGCAGGATTCGTCGTGCGTGAATGAGATCGATAAAATGTTATCGACGAGGAC 60
AGATCGACGATGCGGTTCTTGCCTGTTGAGTGACGCTCCACAACGAGTGTTGCGCCGT 120
GAAAGGCTTGACTIONGGCCGCGACGGACCGCCTTGCAGGAAGGGTCGGTCGGCGATCC 180
CGTCAATAAAATCGAAGCTATTGATCCTGACGGATAAAAGGTCTGCCTCGATCGGAGGAGCT 240
ATTCCGACAGCTGTTACGGCTAGCCATTGGCCGACGGTCTTGCCTCCATTCCCA 300
CGGCGTAGTTAATGCCGGCGGGACGGGAGTGTCTACTATGTGCAAGCACGTCGGCGAAC 360
CATGCCTTCGGATTAATGTGTTCAAGACGGCGGTGTAAGTTGAATGAGTATGACTGCC 420
GCATGGTCAGCGCCCGTGGGAGCCGGCAGATGTCCAGTCGCGCGCTCAAGGCCATC 480
ACATGTTCACTCTGTGGCCAGAAGGCGTCGCTCCTGGTGGCAGGATATATTGTGATGT 540
AAACAGATTAGATATGGACATGCGAAGTCGTTAACGCAATGCTTATCGAATATAAAAT 600
GTAGATGGGCTAATGTGGTTTACGTATGTGAATAAAAGTTCAAGCATTGTTAATAAT 660
ATTTCAATATCGGTGTCTAGAGACCCGTGGATTGTATAGTCAGCACCATGATATGAATC 720
TATAAAATATTGTATCTCAATTGCAATTCAATCGATATAAGAAATTAAATACAAGCCGTT 780
CATATAGTAAGGTTGCCAATGGCATTCAATAACGACCGTACAGTTGCCGCTATATTAAATC 840
TACGTGCCATTCTTAAATAAGATAGGCGAATGACTATCGAAAATAAAACAATTATTAA 900
TGAGTGAAAACGTATTGCACAAATAAGATTCAATTGGTGGCTCAAATTGGCTCTG 960
GTGCTCGATGACGTCGAGATGAGGACAGTAGTGATCAACTGGCGGTGATACCTGGTT 1020
ACGCCACTCCAGAGTGCATGTCGTCCTCCGAGCGGTGAGATAACCCAGTCGGCAAT 1080
TGCTGCTGCATTGCCGGCGTCCCAACCACGACGAATATGCTTCGTTCATCTAACTC 1140
GCGTCGCACTGCCCTCCAGTCATGAAGTCAAAGCCAATTCTACCCCTCTCCATTCC 1200
CAGCTCAGTCGAGAAATCGAACACCTCGTGGCAGCTGACAGTTCAAGAAAGGGCGTAT 1260
CCCTCGAACCTCAGGGTCTTTCACATAGTTAGCAAGGCGTACTGCTGCATAATCTGC 1320
GTTGAAGGCTCTGATGACTACAGGATCCTCGGACAAGCCAATTGATCAGGGCGAACCT 1380
CGCGCTCATATAATGAATTGCGACGACCCCTGCTTCCTGTCGGAGCATCGAATCAATCCA 1440
AGCCTTCCCTGCGGCATAGAGGTATCGACTGCGATGTCATCAAGATCGAGTAGCTTGC 1500
CAACCTAGGAAGTTCTGAGGAAAAATCACCGGCATGACAGCAACCGTCCTCGCCAGTC 1560

AGTTGCCGGACTGGCTCCCTAACGCCATCCACGAATGCCCTACCGCTTGCCTATTGAA 1620
 TGTGTAAAAGAGAAGGACCACTCTTGGCGGTACTTCGGACGCCGGCTAGCCACGCCGC 1680
 AATAATGTGGCCTCAAACCTCACGACCATCCAAAAATATAGTCGCCCTGGATTGACCTC 1740
 GCTGGCCTTGTGAGAAGAGGTTCCAAAAGGGAACGGTGTCTTCGTAAATAGTACTTAA 1800
 ATCTGTGAGTTGCCATGCGAAACCTCTCGAACGATTATCGCGTATCCCTGACATCAGC 1860
 TGAATGAAATTCTGGACGAGTTGTCGGCAAAGTGGAGACCCGCCACGTGTTGAAGTC 1920
 GTGGGAAACGATGGGCACATCGTCGCCGGTGAGTGCAGCTGAGCTCAGAGAGGTTCCG 1980
 CCTGCCAACCTCACCGAGAGCAGCTAACAAACGAAGTTCGGTGCATTCCGTATCCCTT 2040
 ACCCAGATTATACATGCCCGGTGTTGATAACTGAAGAGGCAGTGGCTCCTCAAGATG 2100
 TTCAAGGAGGTGGGTACAGAGTGCAGGGCGAGGACCTCATCCACCGTGACACCAACCGG 2160
 GAGATCCCATTGAGTTCACTGGGCCAGCATGTGCCCGCACGGCAAAGGTTGCG 2220
 CTGGCAAAGAACCCGGCTGCTGCAGGTGGACCTATCCTAACCATGGCAATGGGTTTG 2280
 CTAAAAAGTCAGGCACTTACTGGCAATTGATAGGGTGGATTGCGTTATTAACGTTC 2340
 TCCAGCGGAATCTTATCTTATTGAAATGCTAAAGCACTTAGATAAAATACAGCTGT 2400
 CCGCAATATAAAATAGTAGGATAATGTAATATGTGTATCGAGAATAACGACAAGCTAATAT 2460
 AATCTAGCGTCAAATTGCAATAATTAAATCAAAACTACTGATGAAATAATAAAAGATGG 2520
 TCAATTCTATTGGTAGGAGTTGTCGAAAGATTGACGGACGCCATTACAATACATAGG 2580
 TGCAAGAAGTAAAACAGGAAGGGAAACGGAAAACAGTGCTATAAAAAGCGACAGATCGC 2640
 GGCGATCACTGACTGCGATGGGAAGAAGCTGCCAGTTCACCGAGAATAGCAGAGAGC 2700
 GCATCCTCATGGGTACTACGAACACATTGTCCTACAGAGGGCTTGTTCAGCTGCCA 2760
 ACCCAGAAAGCAAGGCCATTTCAGTTGCCGATGGCGGTGACATGTTGATTGTTG 2820
 CTGCCGTTCCACAAGCGATGTGAAGGCCGATCCGTGAGAGAGGCCCTTGACGAAGGTG 2880
 AAATAGCCTTGGATTTCACACTGTTCAACGGGCACTAGATATTGACCCCTGGCGCG 2940
 GCAACCACCTGAATTGCGAGATGACTGGTTGCCGATGAGCGAAGAAAGCATTCTCCG 3000
 GCTTCTTGTAAAGATTGTGAGATTCCCACATTGACAGCCGTAGAAATGCCCATCGGA 3060
 ATGTTGCCGGATTCCCGGGATGCCACCAAATTGTTCTCCATAGCCCGTGAACGGCITGC 3120

CAGTTGGCAGGGAGAAGAATCGAAGCGATCATCTTGTAGATCGTGACCATTCCATCA 3180
TTTCCCTGGAATCCGATATTTCAATGGCGCTGAAACTGACCTTGCATTTCTCGCAT 3240
TCCCCTGCGGATGTGAGCAATTGATAATGGCCCTTGCAGGCATCCTGGTCAAATTGGCG 3300
ATGATGTTGATGGCAGGATTAATATCCAAACACTGGTGTTCGATCTTGCCTAAAGGTG 3360
GTACCATCGCCGTGAGCAGGGCCGGAGAGATGAATCGGCAAGACTGCGTCGG 3420
ACCCGCTCCGCGGGCGTCGGGAATGAGGCTGATAAGAGACATATCCAAAGGTGTTGTGGG 3480
TAACGGGCTGCTCAATGAAGCCTAAATGCAACGCAACATATGTAAGGATGAGTTGACTT 3540
ATTGGAGAGAGAAATAGGAATGAGCTGCCAGCCATTATCACGTGGGCCATGCTGACA 3600
ATGTTACGTGAAAGGCTCAACTACCTCGAAGCAGACCTCTATATTGTTGACTTTATT 3660
CTGAACAAAGAAGTTGCTTGCACTCATTTCTTAATCTGCCCTTCTGCGCCTCGCTA 3720
TCATGCCGCAACGACGCGACATGCGCTGCCGCGATTGCCCTCCCGAGGGCAACTGGA 3780
AGGAAGAACCTGATGCGCTCCGACCTTGTTGACCCCGTCGAGGTGGTAAGGTGCAAG 3840
TCGGCAGAGGTCTTAGCGGATATGTAATGTTGTCAGCAATGAATCCCACAAAGGTGA 3900
GGGGCCTCGCGATGTCATCGGCAGATGCCGGCTCTTAATCACGTATTGCTGCCGCCG 3960
CCGGCGAAACTCCGGTGCAGACCTTGAATAGGTTACCAAGTGCACATCTGCCACCCCG 4020
ACATAGCCAGTGCAGTGTAGCCACTTCTGAGGGGATCAGCCACGTTCTCCGTGAAAGGA 4080
TTGAGAAAGAAGTTGACCGGGACATTGGAGAAGGGGCCACCGTCTGCATTTGTTGAGC 4140
CGAGAATGAGCTCCAAGGGCTCTCAGTTCTGTCCTTACCCCTCAGTTGCGAGAT 4200
CTGGAACCTTGTGATGCCAGAATGATGGAGAGTTACAATTTCATGAAAGGCAATGGCA 4260
CAGTGACCGCACCGGATTTGAAAGTCATTGGAGAAGCAGCGTATTGACAGGCCAGGCC 4320
CACGTCCGCCACGTCCAAGTTGAACCTCCTCTGCCGCTGTCCCGACAACAGTAAAC 4380
TTGCCGCCACCGATTTACCCATCTGGCCCTGTCGAGCGTGATAAGGAACACTACTCGGCA 4440
GCACGGTATTGGGATTGCCGCTAAGAAACCTGGTACGATGTTATCCGTGCGAAAAGG 4500
TTCTCTGTTGGAGGTGACGTACACGCGCATCGCGCCCTAGAAGTACTTCACCGCCTTG 4560
GGGAACAGGCTTATAGCAATGGCGTGGCACTAGCTTGGCTTCAACCGGTCCGTCT 4620
CTTGCCTTAATCTTCCGCCGCCGCTCGTACATTTCAAACGCTGGATCTCTGTT 4680

CCCTTCCATTGAGTGATGCTTTGTCCTTCTGCACCCGCCACCGCTACAGCGCCAA	4740
GAAAGATGGCCTTCCGATCACTGCCTCTCCCCACGAGCACCAATCAGTCGAACCTCGT	4800
AGAGCCTCAGGTCTCAAGGCATATGTTCTGGACTTTGACGCGCCGACGATGGTTAC	4860
GCCCCGCGACAAAACGCGAGCCAGCTCTGCAGCCAATATGTACGTTCCGTGAACCGCA	4920
TCCCTGTGAAGAGTTCAATGAAATTGGAGTTTGATCCTCGATGCTGCTGCTAAAATGCT	4980
CGAACGTTATGCAAAATTCTAGAAGATGGTGGAAAGAGATGATGATGAAATGGCGAACAT	5040
AATAGATGTATTGGGTTTGTCTTAACTAGTGGATTGATTGAAACAAAGGAGTCCGAGT	5100
TGGGATTCCCTTCCGGCTTCTCGTGTGCAACGATATCGTATGCGTACAGGTATCACATT	5160
AACGTTGCTGCGCGGACCGAGCCCGCTTGGAAAGCGATTGTTGAGCTCCAACTTCTGCT	5220
CAGAAGCACGTGTGGCGAGCGAAGATCATTTGATGAGCAGTGTGGCTCGGGAACGGTC	5280
GCGATCATGGAGGCAACCGTAAATCCAAAACCTGTGTCTGGCGCTGGCAGGAGCGCTTC	5340
ATGACTGAGGGCGTCGATGGCTTTGCACGACAAGAGCAGACCGCCCGCATTGCGCCG	5400
CTTGATGGCGAACTCGTTGAGCGTGTCTGCACTGACGCTTGAGACGCCTCAACAGGAA	5460
GCAACGCACTGGACTGTTCGTGCATGGCCAAGGCCGTTGGGATTGCAGCCTCTCGGTT	5520
GTGAAGATCTGGCACGAGCATGGCTTGCGCCGATCGCTGGCGCTCTTCAAACGTGCG	5580
AACGACAAGGCCTTGCCGAGAAGCTTCACGACGTGTTGGCCTCTACGTCGCCACCG	5640
GCCCATGCCATTGTCCTGTCGATGAGAAGAGCCAGATCCAGGCACACTCGATGGACG	5700
CAACCAGGACTCCCTTGAAGAAAGGGCGCCGGCACAAATGACCCACGATTACAAGCGC	5760
CACGGCACCACCCATTGCGCCCTAACATCCTCGACGGCTCGGTGATCGGCCGA	5820
AACATGCAGCGTACCGGCATCAGGAGTTCATCCGTTTCTCAACGCCATCGAGGGGGAA	5880
CTGCCAAAGGACAAGGCCGTCACGTATTCTCGACAATTACGCGACCCATAAGCAGCCG	5940
AAGGTCCGCGCCTGGCTGGCAAGGCATCCGCGCTGGACCTTCCACTTCGTCCTAACATCA	6000
TGTTCATGGCTGAACGCCGTCGAGGGATTCTCGCTAAATTGACACGTGACGTCTGAAG	6060
CACGGTGTCTTCATTCCGTCGTTGACCTCCAGGCCACCATCAACCGCTCGTCAGAGAG	6120
CATAATCAGGAACCAAAGCCGTTCATCTGGAGAGCAGATCCAGACGAGATCATTGAGCC	6180
GTCAAACGTGGCACCAAGCGTTGGAATCAATCCACTAGCGTATGAACAGTAATAAGAAA	6240

FIGURE 2, Sheet 5

ATCCCGATTGTGAATAGTCCCAATTCAAATGTGTCCGTGTGAATTGCGTGTCTTCAG	6300
TTGAATTCTTTAATAATATCAAATATTCAATTGTGAAAAGTTGTATTGGTTCAGGTTC	6360
AAGCTTCCGAATTGTTGAATTTATTCCCTGTTCAATTGTTGACTGTTGGGAG	6420
ACACCTTTTGTGTTCGTAACATGTCACCCCTCGGTATACATTAGCCTACAAAGTA	6480
AATAACGTTGATAAAATGTCACTCATGTTGTAATAAAATTGAGCTTATTATGTATAACCAG	6540
ACCTGTGTTAATCTAATTACAAAGAAATTCACTCATTCTCCAAAGCAATCCTGAGTAGCT	6600
GCGTGATGGATCTTCATATCAGCGCCCACGTTCACCCCGTTGCCGTACCCATCCAC	6660
GTAGTGGAGTCAACCTGAACCGTGCAATTCTCAGGCCCTGTCTGCTATGATCAGTTCT	6720
GCGAACGGCTTGCGATATCAGCAAAGCTGGACGGATTGGGTGTTGACCCACGGATTG	6780
CAGAAGCCATTGAAGACGTGGCGCTGGTGTCCAGGTTGCACCTTGCCTTCATGGCCCCC	6840
GAATAGGCGCGCTCGAAGTGTGATACCTCGTCGACCCAGGTCTTCATTATATGTCGA	6900
ACAACCAATTGCAGCGCTTGTGACACCCAGTGCTCAACTTGGCGACGCCGTGC	6960
TTGCTTGCATGATCCGCCCTACCGCGAGTGACCTCTCGCTGCAGGAAATGGCTGGCGC	7020
ACAACAGATTGCCAGGCAGTTACACGAGGTCCGCAGACGTACAGTGCTTATGCCA	7080
TCCAACCTCAGCAGCCGATTGTTGAGGAGGGCACATGTAACGTGCACGGCGAAATGGCT	7140
TAAAAAGAACCTGCCGTTCTTCGTCGCCCTGCTGAGTCTTCAGCCGTTATGACATCG	7200
TTGCCATTGGGCCGGTGCTTCCATGATGAACTGGATTGCCAGCAAATGCAATGAGC	7260
CTCTTCTGCTTGCACCTCGGGTACGACTATCAGGTTTCCCTCCAGGAGTGCATGCC	7320
ATGATGGTGTGGGCATTATCCGGAAGGCGCACCACTACCTAGTGTTGCCATCGTAGGAG	7380
GCGGGCTGCTGGCCTGTTGCTGCCACAGAACTACTTGGCGCTGGCGTAAGGAAATCA	7440
CTCTTCTGATACCGTTGATGAGATCCGTAGTTGGGGCATGCCGATGCCAACGGCG	7500
ACGCTCACCAGGCCCTGACGTGTTGGTGTATGCCCTCTCCGCCAACCAACTTGGC	7560
TGTCATACTATCTGGATAAGTTAGAATTCCGTCCAGCCTTCGTTTCCCTGTGCCGGCA	7620
ACGACCACACAGCACTATATCCGCCAGAAACGCTACGCATGGCACGCCGGGCAAGCTC	7680
CGCCGGGGATATTCAGCGGGTACATGTCGGATGGAAGACACTACTCTACCAAGGGTGTG	7740
AACGGAATGGCAGGAGACTGATGGCTCCGATGGATATCTCTTATGTTGAAAGAGCGTC	7800

GTCGTGATGAAGCCTCAGAAGCACGGCAGCTTGGCTCCGAGAGTTGGAAAATTCACTT 7860
 TCCATGCCGTTTGGTCGAGATCTCAGCTGTGGAATTGAGTCCTGGTGGCAAGGCAT 7920
 GGCAAAACACCCCATGATTGAGGCTTCGGGATACTGAGGTTGGGATAACGGCCGAGTT 7980
 CGTCCTATTACAACGTGTTCAACGATCCTGGACTGGATTATCAATGGCTACGAGG 8040
 AGGACCAGCATTCTCTATTGGTGGGGTCAACTTTGCAGGCTCTGATGCGCATTGAAA 8100
 TATTCCAGAAAAGCCATGCGAAAGCACGACTCTGTTTGATCCC GTGC GTGGAATAGCC 8160
 AGGAGGGCGGGAGATTGAAGGTATGCTTGAAACACGGTCATTGCGTGTGACCGAGG 8220
 TCATCATTGGCGGCAGTGCTGAGGCCGCTACAGTTGATAACAGACTGGCCGGGGATGAGA 8280
 CTTCCCTCAGCTACAATATCGAACCCGCCGTCGGAAACTCGTCTGCCGCTGTCAATTCA 8340
 CACTCTTCATGGTACGAAGCAAAAGTTTGGGTTAACTCCGGCATCCCAGCAGTGATAT 8400
 GGACCGATGGGCTTGTCCGTGAGCTGTGTTGCATTGACATCGAACGCCAGCTGGAGAGG 8460
 GCCTTGTGTTTCACTATGCTTGATGACTATCTATCCGGCCGATCGAGCATCATG 8520
 ACAAGAAGGGACGGTGCTTGGATTGGTCAGGGAGCTTGCTGCTGCCCTTCCCTGAACCTGG 8580
 CTTGTCACCTGGTCCAGTCAACGAAGACTACGAACGATATGTCTTCGACGACCACCTAA 8640
 CGGATGGTTTAAGGGAGCTTGTGGAGGGAAAATTCTCTGGAAAAGGTCAGTATATCC 8700
 AGGATCTGCCCTGGGATAATTTCCTATTGGGATCACGGGGAGCCTATCTGATTGACC 8760
 GTGACGACTGCGTCACCGGAGCCTCGTTGAGGAGCAGGTGAAGGCAGGCAATCAAAGCGG 8820
 CCTGCGCCGTATCCGCAAGCACCGGCGGGACGCTCTCTTCACTCCAACCGGTGGACTGG 8880
 ATAAAAAAATAGAAATTTCCTGATTAAGTTAGTCAATGTAATTGCGTGTAAATCCG 8940
 TAGGTATGCAAGCTGCACCGGCAGCATCATAATTGATGTTCCATCAATAAATTAAAGGTG 9000
 CCCGTTCATGTGATTACATTATGTATGTTATCAAAAATATAATCGAAGTCCATTAA 9060
 AGTCTGATATTAAATTGGAATTCCAAACGATTCTTGATGCCATCTCGTATGATTGTA 9120
 TGGTAATAAAGTCTCCACATCTCCGAAAATGCTTGTGATTTACTTGTCTCTCACGT 9180
 GCTTTCGCACTTGACAGCCAAAGTGGCAACTGAGAAGAGTATTAACCTGGCACGCA 9240
 ACTCGAGATATTCCCACTAACCCCAATGACGTATTGCACTCGTCACGGGTAGCAGCCCC 9300
 ACTTGCCTTGGCCACTTTATTAAATTCTTGGCCACTGGCATTAAATTGGCACCTACATA 9360

TATTAGTGGAGAAGATAAAAGTGTCACTATCGTTCTGTCATTTGAATTTGCAAGG 9420
 ATTCATGTTGTCAACTACACAGCTGAAAGGAAATCCGCAATCAACGGAGAACGTCAA 9480
 CATCTCGACAAAAAAAGAATGCTTCATCATTCGCTAGACTGCATATTGACCGCTCCTTC 9540
 GGCGCTGGGCTGCTTTACTGTTGCCAGCGTCGGACAGCCACCAAGAGAATGGGCTAT 9600
 ATAGATCCTTCATCAAACCAAAACATTACTAAGATCATGCTGTAACGCTCAATACGGT 9660
 GAGTGTGGTTGAGGTTCAATTATTACTATTTGAAGCTGTGTATTCCTTTCTAA 9720
 TATGCACCTATTCATGTTCAAGATGGAATTAGCCGGACTAAACGTCGCCGGATGGCC 9780
 CAGACCTTCGGAGTATTATCGCTCGTCTGTTCAAGCTGTTAGGCCTGCAAAGGCCAAG 9840
 AGGAAGGCCAAACGGGTATCCCCGGCGAACGCGACCATCTGCTGAGCCAGCCAATCTG 9900
 AGCACCACTCCTTGGCCATGACTCCAAAGCCCACCGGGACGTTCAACGACCCGCGAG 9960
 TTGCTGCGAAGGGACCCCTTGTGCCGGACGTGAAAATTAGACCTACGGGATTAATACG 10020
 CATTGAAACAAACCTACGGGATTAATACGCACGTGGCTGGCGTCTTCGATTCAATTTC 10080
 CACGCCGGAGATGATATCGAATATGTTCTGTTAAGTTAAAATAAGCTGCGAGCCATGGCG 10140
 CGATTGTCCTGTTTATTAATATAGTACTTTAACGTCTCTTAGAGCGTTGTAAATGT 10200
 CGTGAAAATGTTTATGTCATGAAATGTACTGTTGAACCTATAATATTATAAGTCCAGGTGT 10260
 CGTTGTTGTTGATACTGCAATATATGTGTAGTAGATTAGATAGTCATATGAGCATGTGCT 10320
 GTTTTGGCAAAATTAGCAGCAGGATCAACACAGAAGAAAATTTAGTACAAGAAAAT 10380
 AGGTCAACACATTACAACGTACGCTACAACCTCCAAAGGTCTGTGTCACAGACTGCGGG 10440
 GGGTACATAGAACTTATGACAAACTCATAGATAAAGGTTGCCCTGCAGGGGGAGTTCAAGT 10500
 CGGCTTCTGGCTTCTTCAGGTTACTGCAGCAGGCTTCAATGACGCCCTCTCGCT 10560
 TCCTGATCAGGCCCGAGAGTCGCAGGGTAGGTCTGGCTCCGGTGAGGAGGCGGCCGG 10620
 CGTGATATCCCGAGGGCATTTGGTGAATTGTGTGGTCCGCAAGCTACAACATCATAG 10680
 GGGCGGTTTCAGTCCTCGCCGCAGAAAGAAGGTGCAAGCTACCTCTCTCCGTAAACG 10740
 TTGGTCACTTTAACTCCAGCAAGTGAATGAAACAAGGAACTTGCAGAAATGGCGATGAAG 10800
 CATTCTAAATCAGGTTCTCCGTGCGCTGTGCGGCCAGCAAGGTTGTGAACACGGAGC 10860
 ATCTCCTGGAGGGCGAGCTGCTCCGATATGGTTGAATCGTGTGCCAGCACGGCTCC 10920

ATTCCAAATGTAATGGATTGTTCTTCAGCACTTCTGCATCTTCGCGAGAAAGATAG 10980
 ACAAAATACATGTTGGTCGTTCTCGAGCCAGATCCGGCTGACTAACAAACATAGGAGGA 11040
 TGATAGCAGACTTGTCTCAAGAGCTCAGCTAGTTGTTAAGTATATATATCGGTGG 11100
 GAGTTTCTTCAAATCTAGCACTGCAAGAGCCCATCGTTCTGGAAATGCAGGAGGGT 11160
 TTGCTATAGTCACGGCTATAGATTGCAAAAGCAAATCGGATCCCCTCGAATAGGTTATC 11220
 TGGCTCCATGCTGGAGTGAGATCTACTGGTTGAAATCGTGGAAAGGAATAGCAATTGG 11280
 TCCATTGTGATGTGAGTTGGATAGTTACGAAAAAGGCAAGTGCCAGGGCCATTAAAAT 11340
 CGGCGTCGGAAACTGGCGCAATCAGACACAGTCTCTGGTCGGAAAGCCAGAGGTAGTT 11400
 TGGCAACAATCACATCAAGATCGATGCGCAAGACACGGGAGGCCTAAAATCTGGATCAA 11460
 GCGAAAATACTGCATGCGTGATCGTTCATGGGTTCATAGTACTGGGTTGCTTTCTTG 11520
 TCGTGTGTTGGCCTTAGCGAAAGGATGTCAAAAAAGGATGCCATAATTGGGAGGAGT 11580
 GGGGTAAAGCTTAAAGTTGGCCCGCTATTGGATTTCGCGAAAGCGGCATTGGCAAACGTG 11640
 AAGATTGCTGCATTCAAGATACTTTCTATTTCTGGTTAAGATGTAAAGTATTGCCAC 11700
 AATCATATTAAATTACTAACATTGTATATGTAATATAGTGCAGGAAATTATCTATGCCAAA 11760
 TGATGTATTAAATAATAGCAATAATAATATGTGTTAATCTTTCAATCGGAATACGTT 11820
 AAGCGATTATCGTGTGAATAAAATTATTCCAAAAGGAAATACATGGTTTGGAGAACCTG 11880
 CTATAGATATATGCCAAATTACACTAGTTAGTGGTGCAAAACTATTATCTCTGTTTC 11940
 TGAGTTAATAAAAATAAAAGCAGGGCGAATAGCAGTTAGCCTAAGAAGGAATGGT 12000
 GCCATGTACGTGCTTTAAGAGACCCCTATAATAAATTGCCAGCTGTGTTGCTTGGTGC 12060
 GACAGGGCTAACGTGGGGTTAGCTTGACAAAGTAGCGCCTTCCGCAGCATAAATAAAG 12120
 GTAGGGGGTGCCTCCATTATTAAAGGAAAAAGCAAAAGCTGAGATTCCATAGACCAC 12180
 AACCAACCATTATTGGAGGACAGAACCTATTCCCTCACGTGGTCGCTAGCTTAAACCTA 12240
 ATAAGTAAAACAATTAAAAGCAGGCAGGTGTCCTTCTATATTGCACAAACGAGGCGAC 12300
 GTGGAGCATCGACAGCCGCATCCATTAAATTAAATTGTGGACCTATAACCTAACTCAA 12360
 ATATTTTATTATTGCTCCAATACGCTAACAGAGCTCTGGATTATAAAATAGTTGGATGCT 12420
 TCGAGTTATGGGTACAAGCAGGACCTGTTCTACTTTGTTAACATGGCTGAAGACGACCTG 12480

TGTTCTCTCTTTCAAGCTCAAAGTGGAGGATGTGACAAGCAGCGATGAGCTAGCTAGA 12540
CACATGAAGAACGCCCTCAAATGAGCGTAAACCCCTGATCGAGCCGGTGAGAATCAATCG 12600
ATGGATATTGACGAAGAAGGGAGGGTCGGTGGGCCACGGGCTGCTGTACCTCTACGTCGAC 12660
TGCCCGACGATGATGCTCTGCTCTATGGAGGGTCTTGCTTACAATTGGATGCAAGGC 12720
GCACTCCTCACCAACCTTCCCCGTACCAAGCATGATGTGACTCTCGATGAGGTCAATAGA 12780
GGGCTCAGGCAAGCATCAGGTTTTCGGTACCGGATCCTATGCGGAGCGCCTACTTC 12840
GCTGCATTTCTTCCCTGGCGTGTCAAGCTGAATGAGCAGATGGAGCTAACTTCG 12900
ACAAAGGGAAAGTGTCTGACATTGACCTCTATGCCAGCACCCAGCTTAGGTTGGAACCT 12960
GGTGAGTTGGTGAGGCATGGCGAGTGCAAGTTGCAATCGGCTAATGGTAGTCGATGGG 13020
CTGACGAGTTGATGTCAGGAGAAGCTGAGTGTGTCATTGTTCCCTTAAGAAGTATT 13080
AATGTAATAAAATCAAGATCTGGTTAATAACTGGATACTTGATTCATCGCGCTTTT 13140
TTGAATAAAATGTTGTTGCTTGACTTTAAGATATCCTTGAATTTGCGTTATTGCTAT 13200
TTCGCTTTGGTTATTCCAAAAGACTTGCTCAGTAAGATCAAACGTTGTATTCTCC 13260
GGGCCACAATATTGACCTATATGCACTGGCCCACGCCGCAATAGATGAAAATTGCCA 13320
AAATTAGCTATCGGTCTCTGAAAAGAAGGGCCGACATGTTTCTAGACCATGCAAAGT 13380
CATACTACCTGAAACTGATAAAACGACAAAGAAAGTAGCCTATTAAAAGTCGCTATA 13440
GCATGAATTCAACACAAGGAAACCAAAAGTCGGAAGGAAGACTTTAATCCGGATTATT 13500
GGACATGATAGGAGCTATGGGGCAACGTGTCTTCAATGAGTGTGAATGATTTCTGT 13560
AGCAAATAGAAAACGTTTTAAAACGATGTGGCCTGGAGTAATCAGCGGAAGAAAATGG 13620
TCATGCTCAGATAATTCCGTTGCTGACCTCGCAACCAACCCCTTAAATACCTCTGCTG 13680
CCCATGCATTTGCCAAGTTAACCTAAAGTGGCAGCTGAATGGCTCGTTATTGCAGTGGT 13740
GGCTCTCAACGGCTTCATGTCGATGATTTCTGTTGGATCAAGGAGCCCACTCGACTGAAG 13800
GCTCAGCTTATTAATGTGGGGAGACCTACAAGGCTGCAAAACAGAGACGTTAAAGTAC 13860
TATATATCACTGCAACTGAGCGTGTGGCTCATGTGGAGGCAGCCGAGGTCAACAATGCG 13920
GAAATGGAGCTGCATCCTGCTGGGTTGAAGTACCCCTCTGTCCTCGTCTTACCTCCCTG 13980
GCCGTGGCTACAGCCTGCAAGGAGAACAGCATCTTGTGCGAGGAGCATTGGAGGG 14040

GACTTGATAATCGTGCCTCGTTCCCTCATCAGACAAATGTCTCACTCGCTGCTTAAGG 14100
 GAGCTCCACAAATTCCATTTCGGGAGGGAGGGTACCAAGGAACAAGCAGACATGGATTATTT 14160
 GTGGCGATCATCCAAATGATAATTCTGACTATCAGAGCTGCCTGAAATCGACACACGAAGT 14220
 TGCCTAAAGGACTTTGCAAGATTATAGTAGGAACTGGGAGGGCAGCCTCTAGCTTAT 14280
 GACGCCATACGGCAATCGGCAAGGTGCTGCTGGAATAGATAGTGGGCCCTGATCC 14340
 GAGTTGATTTCGTTACGTAACTTTATCATGCATGTTGCTTATGC 14400
 TCCCGAGTGTGGCCATGTTGTTAAAATAAAAGGCTGATGTTAAGTCCATTGTA 14460
 AAATACCTTATAGATTAATATATAGTATAACTCTGTATGCCGTCGATGAGCGGTT 14520
 ATATGATTGAAATCTATACGTTGCAATCAATCGTATTACAGTGAGCCGTGCTTAATG 14580
 AAATAAACATCATGTTAAATGTCTATTATCAATCAACATGCGCTGACAATAATCAAA 14640
 GGGGAAACGTAATAACATTGGGGATAACAGCGTTATTGGGAGGTCCGGGGCCGATA 14700
 CACTTAAATAACATAGACAGAATTGAGAGAGCACGCAGGTTGAGCCAAGTTGAGCGAC 14760
 TTGCCGGTAGCACGGAAGCTAACGCTCAGGTGTTACAAATAGACAGGCGTCGAGGCACGA 14820
 GCACGACGACCTTGCCGGACATTGGCTCGCAGGGGGCTCAAAGCGGTTGGCTGTAACG 14880
 GACCTTGTGTTCTTGTAGCTTCATCGAGCATAACCATTGGGACGGTTGCTGAACA 14940
 ACGGTAACGCACTTTTCAAGGGAGCGAGGTAGAAGAACATATTCCCCGTGGCAGCC 15000
 GGCGGTGAGCATGCCATTCTAAGGGATCAATGGACTCGTGCACGGTGAGCATGCC 15060
 TTCTGACCGTCGGTGCCAATCAGCAGGCCACTCCAAACATGTTTCAAGTCCTTAA 15120
 CCAGTCTTATAGCATTGATCTCCAGCAATCTTATTGAAAGTCGATTTAATATTCAA 15180
 AGAAGATTTAGTGGAAAGGGAAATAATCGCGTGGCCGAAGAAGAGGCCCTCAAAATCA 15240
 GAATCCACTAGGATAAACATAATATCTGAAAAGCATTGAAATTGGGTTAGGCACGAGAG 15300
 GCTGACGCGGATGCCACTCGATTGCTAGTGGAGGGATTCCCTTTCTAGCGTATCGAA 15360
 TTCACCGTTCACTATATGTTTCTGATTGGTTGATCTGCCGGGACCACTTGACTGCC 15420
 ACTAATATCGAAAGTGGCTGCTCGATTATGATGCTTGTGAGAGGTTCTCTCCAA 15480
 TGCATGCAAGCTGGCAGATCGGATACTCTCAATAGAGATCTTATTGCGCTCAAA 15540
 GTTCCCAGAAATCAACAAAGGGAGGGCAGGTCTTAAATACGTTGCAAGCTGTCCTTAA 15600

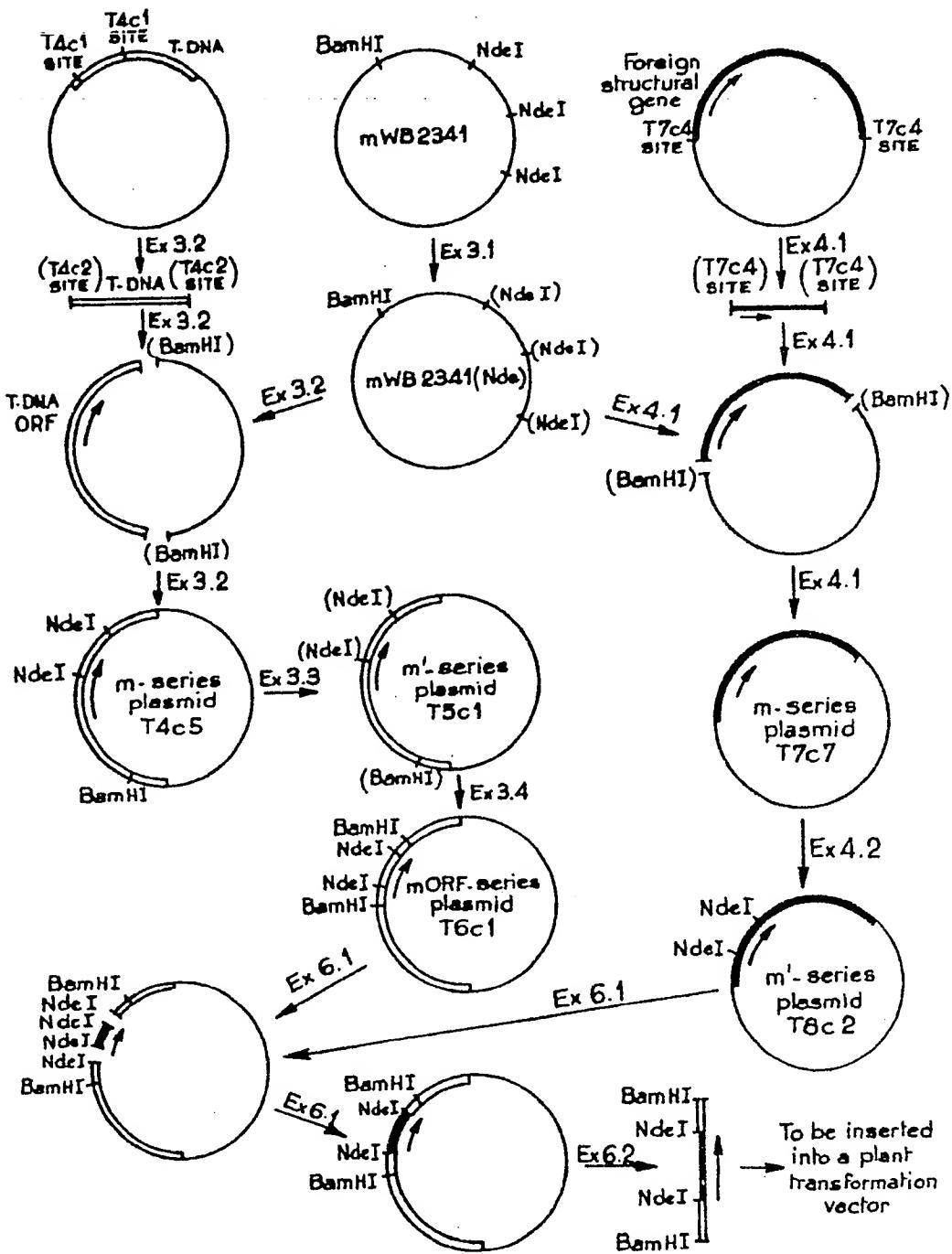
AAATAGAAGAGAATTACAGCTGGAGGCACAGACCACTAAACTGCGAAAGTAAGCATGGC 15660
 AGATGAGTTGGAGCGTCAATTGGAAGCCATTCTCTCATTACAGTCCTGGTCCGGATGT 15720
 GAAGGCTGAGCTTGAGGCAGCTACGAGACTACTGCGAAGATCTGACTTCTGGAAAAG 15780
 CCACGGTTTACCGGTGGCGGATCTCGATCAGACTGTGACTGTGACAAAGCTTCTATACAT 15840
 GTATATGGATCGGGCAACAGCAGACCTGTGTGAAGAATCGCTGCCTCGTTGCAACAG 15900
 TGGCAATTCAAGCCGCAAAAGTAACCTCGCTTCCACCATAACCTTGCAAGGCGTGACAAGCGC 15960
 CGAGGCCTATGAGAAACTCAACTCCATTGTTGATGGGAGTGTGCGCCCCCAATCTCGTGG 16020
 GCCTCCCTGCTATTTGTGGCGTTCTGCCAGCAGCTGTTGAGAAAACCAGTGAGAT 16080
 ATCGGTGCGCACAGTGGACGGCGAGTGTGGCCCTTCGATGTTTACCCGGCAGCGTC 16140
 GCCACAGGATCAGAGTGATATGTTTAAATATGAAGGAGTTGATGTCGGAAAGAG 16200
 TGTATTTATGTAAGAATTATCTTTATAGCCTGTGTTACGTTGAACCCGGTCCGCGCGG 16260
 TATTGTTTCAATAAAATGGTATGTCGGAGGGATATAATTGGTCTTCATTGGTGTGATT 16320
 ACGTGTAAACGCGGATAATAATAAGTAAATTACAAAAGAGAAACGCATAATTTATTCCA 16380
 GAATGATTGCGAGAACGATGAAAATACATGAAAATGCATATTGTCGCCAGGGAAAGGATG 16440
 GCGCCGAAATAACGAAACTGAGCCAATACAGTGACTTGCCAAGCGAGTTGATCCTACC 16500
 AAATTGCGCAAATTAAATGCCGTGTTCCATGGGCCAGCGAGTTATTCAAAAGAGTT 16560
 CGTACACGTGGCGGCACGGCAACGTCAATGCTTAGCCCTACCGGGAGAAGTTGG 16620
 CGGCCCTTCATGCCCTGAGGTCAATTGTCAGGCTCGTCATCGAGAATTTCGGTGT 16680
 AGTTCTTGATCCCATCGCGCTTGCCGTGTTGGGTCAAGTTTCAACCGCCCTAGAATAGT 16740
 AGAGGGCAACGGCATCAACGTTGGGGCTTCCATCGCAACAAGGTCACTGGCGACAATT 16800
 GACCATCCGAGATAGGACATGCTCAATGTAATCCGGCGCATGTCATCAATACCGAGTG 16860
 ACAAAAGTGACTGCGTTGGGGCGATTTCAAGCGCTTCAAAATCCGGTTTCCGTAGTTGG 16920
 TCGCCATGATGACGAATTGAGAATATGGCAAAAGGCTACGATGCCGACAGCTCAAGGC 16980
 TAAAGGTTACGCAATCACGTAACCTTCGACGAGCTCGAAATTGGATTCTTACCGCCGC 17040
 TGAGCACTGCTACCTTACGAATTCTCTTAGCGGCACCATAGTTAAGTGAGAGAAATTACAG 17100
 CTTCGGCAACTTTCCAGCCCCAAACAAGAAAACGTCGAATGTCCTCTGCCTTGCAACAC 17160

GCAGGTTACGCATGCTAGCGAGAACCAACCGTTCTCCATTAGAAATTGCCACGCCCT 17220
CTACCGACATAAGGAGCGTCCCGGACACCTTGTGCGCAGGAAAATATCGGAGTGCTGG 17280
GCGGCTTCCGGTAGCGGCCTGGTGGCGCGAAGTGGATGTCTTGGTGCCGGAAATATC 17340
TTCCGAAATAGCCAATGAGTGCTCCTCAGCCATCCAGGAACATTCTTGGTAAACGTTA 17400
GGTAAGCTTGTACATGTCGGCTTCTGCGGAAACACCTCCAATAGGACTTGAGAG 17460
CTTCGTCAACAAATGCTGGTGTGATCTGGATATCGAGGTGTGATAGTGCAGATTCACT 17520
AGTGTACCTCGCAAAGTTGTTGGCCATCTGCCTGTAGGTGCGAATTTCCTGCTCAA 17580
ATTGTTGAGGTTAGCGGATTGTAAACGCGTTATATGGGCTGCTGGAGGGTACTTTG 17640
GATTAATTTCTGCCAGCGCATTCTGACGCGGCACCGCTTGGAAAGTGCCTGTGG 17700
GTCCCGTTCTACAATAATGTGCCGATCCGGTCAGAAAGTATATGGATGAGTTGTGCC 17760
AGCCTCACCAACGTGCTGCAGGCCATCATGACTACTCAATGTTAATGGGGTAATGAA 17820
TAAATAGGCAGAAATTGGGTACGGTGGGCCAGGGAAATATAATATTGCCAGAGGTAG 17880
TCGGATGCCAAGGCCGCAACTAATAGTCACGAACAAATTGATGGCTAGTGGCGGCCAA 17940
CTCCAAAACCAATTGCCAGTTATTGTATTGCAATACATATGAGTATTGGATACAAC 18000
AATTTCATTAATAATATTTAAGTGTGGACAGAATAGCGCCTAATAAATTGCGAATGT 18060
TGTCCAATTGACGTTTATAGGTAACTCGATAAAATCGTGTCTTGTGATATTCTGATGC 18120
GGACAATATACATTTAACATAAAGATATAAGTTATTGAGGCATTATGATATTACAAT 18180
AGTGGGGTACATTTTACAGATGCTGTCACCCATGAAATATTGGCAAAATACTCTTAA 18240
ATATGCAAGAAACTAAAGAGGATGCATGGTTGGCTGTAGGTACATGGATGCAAATGCT 18300
GTTTGCAATAAGTCATATAGTCGTCTGTTGAGTGAGGCCATTCAATCAGCAAGTAG 18360
GAATGAGGTGCATGACATATTTGAACCAACAGTTGGCAAGTTTCTACACAAA 18420
TGCACGGCTACGGCCAAATCGTAGCTGCAAGTCCAATGCTGAAAGTTAGCCGGCCCG 18480
TTCCAAGAAATTAGCCTTGCATAAGGACTGGATCGCGGAGAACCTCAGAGTAGTTCTG 18540
ATCCCATTGTCCTGCCGTGTTGTTAGCTTAAATGGCGTCTGAAATAGTGCAGGCC 18600
AACGAGTCGATATTACGTGTTCCATCGCATCCATATCATGCCACCAACGATGCCACTC 18660
AGCTCAACACGTGATCAAAATAGTCAGCTGGCAATTGTCATTCCAAGCGTCAATGTA 18720

ACGGCATTGCTGTGATCTCCTTCAATCTCAAAGACGGGCTTGTGTTGAATTCTCGCGCGTA 18780
 ATTATGAACCTGGATTTGCTGAGATATGCTCGATTGTTAACAGCCTTGAGTGAAATCTG 18840
 ACTTCCGGCTGAAGCCTTGCACCAACTCATGGTTGACTGGTTGCAGCGGCTGAGAAC 18900
 GCGATTCGTTGAATTCTTCAGATGCTCCGAATTGAGGGCGAGGATGATGGCCTCGGCA 18960
 ACTTTACCTGCTCCGAATAGGAAGACATTGATCTGGCTCGGCCCTGCAATAGGAGATT 19020
 AGGCATGCTAGTGCCAGCCAAACAGTTCTCTCTCGATATGCCACCCATCAACAGAG 19080
 AAGAGACGTCTACCTGTGAAACGATTGCGAAGCCAACGTCGATGTGAGAAGTCGGTTCT 19140
 TGTATCTCGCGTTGACGGATTAGAATGGATGCTTTCACACCCGAATAGTCGCCGACGA 19200
 AACCCACCAAGAGCTCCCTCCGTACAGCCCTCTCGATCAAGTGGAACGAAGACCTTGTG 19260
 GGCGGAGCCGCCCTCAGCAAAGAGGTGCCAATAATCTTCAAGGCATCCGCGACGAGTT 19320
 CCGGTGTAATGTATATTCAAAGCCGATAGAGATTCCCTGTGTCACATTGCTCGTGT 19380
 TTTGATCGGCCATGTTGTGTTGATCAGCCTCCTTCGAAAATTCTTGAGTTCGAAT 19440
 AATTCTAAAATCGAAGGACGATTAATAGTGCCATACCAAGACAAGAAGGTAGGTGGGCC 19500
 ATCAATCCACAAGCCTAGCACATTGCTGCTGCTCATGCAAGGTATCCAATGGAAGCC 19560
 TGGATTGGTTAGCCGAACCTGGTGGTTCAATTGGAGCGGGCAGGTACCTTTGTCTC 19620
 CAAATAACTGAAACTAAGTTTGTATTGGTATGTGTTGTCTGTTCTGCCGAAGGTGC 19680
 CCGAATTGCGCAAATTCCCTTCTAAAAGGCTTACATCTAGCAAAGGTGAGCCCTGTG 19740
 CATCCCAGCATTGGACAAAGCGGCCAATTGGACAGCGACTGGCTGCGTTGGAGGCTC 19800
 GGATCTCAAAGAATAGAAAAGAGTTATGATCATGTTAGAACCGCCAATTGGCGGT 19860
 TGAGCTCTTGATGAAAGTAATGGTTCAAAAAGCAACATCGTGGGTGAAAGGTACCTA 19920
 CATATCTCACAGACAATAACTACTGTTGCTGTTGCTGATTGACTGACAGGATATGT 19980
 TCCTGTCATGTTGTTCAATTGTTCAATTGTTCAATTGTTCAATTGTTCAATTGTTAATG 20040
 TATAAGTTGATGAAGGATGGTTTTAAAAATAGTATGTTGACTGAGGTAAAGTC 20100
 ACTCACGTTTGACATCGACGGACCGTAAGCATTCTTGGTAAGACCGAACGCTCGTCC 20160
 CAGATAATAGGCCCGTGGAGGGAGGCCTTGTATGGGCCACCGATGGCGTGCTGAGCC 20220
 GAGTACGGCGACGCCTGCGCGATTGCGCGGGCGGACTGCGCGAGGGCACGGG1TC 20280

TACGAGGACGAGCGTAAGGGGATAGAGCTTCCGCCGTCGGGTTCAGCCATATTGCT 20340
TGATTGCGGCCACTGGAAATGCAGCCAGGTCGTGCTGCCGGCGGCCGGTCAGCGG 20400
CATGCTGCGCACCTCAGTGTTCGGCTTCCCTCAGCTCACGGTTACTGTGTGGCGCTAAGA 20460
ACCGAGGCCCTTGATGGCGGACCTGGCTTTCAATCAAGGGATGCTGATTTCCATCCG 20520
TAAGCGTCTCGACGGTGGTACACCGTCGGCTCGGCGACGATGAGAACGAAATCGT 20580
CAGCGATAGCTTCGTTTCTGTCGGATTATTCCTCCTGATGCGAGAGGAATGGCTTC 20640
TATCCGGCTGCCGGCGGGTGCACGGCTCCGGATTGACCCCTCCCGTTGCTTGCTACTTGGC 20700
TCGAGTGACGAAATAGCACGCCCTGTGCCGCTGTATCATGTCATCGGGCTCACAGGAGAT 20760
TCGCTCGTAGCGCGTTGGTGTCACTCACCAACACCGCGTCGTCGACCAAATTGGGGAGGA 20820
TGGTAGCGGAATCCTAAAATCCTAAAACCATACCGACCGTCACGGCGCTCGTACCCCT 20880
GCGAGCGACGCCGCACTCTCACCTGATCCGTGCTGCCGGTGTCAATACGCAATGAGC 20940
ATTGTCACGGTTCTCAGGGTAAACGGCAATCTTCGTCAATGCCGGCGTGGATGCTATCA 21000
CCGTTAGAAAGGGCCTGCCCTGCCCCATGGTGGGTCTCTAAGGTTCAAGTCTGAGAACGGCAGC 21060
CAGAGCGGCACTGTTGAAGAGCAGTCTGAACCGCTCAGATCGCTCGCATCGATGCTTGG 21120
GCGGCG 21126

FIG. 3 Schematic Diagram of DNA Manipulations





Europäisches Patentamt
European Patent Office
Office européen des brevets

⑪ Publication number:

0 204 590
A3

⑫

EUROPEAN PATENT APPLICATION

⑬ Application number: 86400855.2

⑮ Int. Cl.³: C 12 N 15/00
A 01 H 1/00

⑭ Date of filing: 21.04.86

⑩ Priority: 22.04.85 US 725368

⑪ Date of publication of application:
10.12.86 Bulletin 86/50

⑫ Date of deferred publication of search report: 21.09.88

⑬ Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

⑭ Applicant: INSTITUT NATIONAL DE LA RECHERCHE
AGRONOMIQUE (INRA)
149, rue de Grenelle
F-75341 Paris Cedex 07(FR)

⑮ Inventor: Slightom, Jerry L.
3305 Lorraine Avenue
Kalamazoo Michigan 49008(US)

⑯ Inventor: Tepfer, David A.
9 Passage de la Main d'Or
F-75011 Paris(FR)

⑰ Representative: Warcoin, Jacques et al,
Cabinet Régimbeau 26, avenue Kléber
F-75116 Paris(FR)

⑲ T-DNA promoters of the Ri plasmid.

⑳ The sequence of the T_L-DNA of Ri plasmids found in *Agrobacterium rhizogenes* strains HRI and A4 is disclosed. Sixteen open reading frames bounded by eukaryotic promoters, ribosome binding sites, and polyadenylation sites were found, five of which were observed to be transcribed in a developmentally and phenotypically regulated manner. The use of promoters and polyadenylation sites from pRi T_L-DNA to control expression of heterologous foreign structural genes is taught, using as examples the structural genes for *Phaseolus vulgaris* storage protein (phaseolin), *P. vulgaris* lectin, a sweet protein (thaumatin), and *Bacillus thuringiensis* crystal protein. Vectors useful for manipulation of sequences of the structural genes and T-DNA are also provided.

EP 0 204 590 A3



DOCUMENTS CONSIDERED TO BE RELEVANT					
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)		
Y	INT. REV. OF CYTOLOGY, suppl. 13, pages 105-125; J.M. JAYNES et al.: "The position of <i>Agrobacterium rhizogenes</i> " * Page 120, lines 8-12; figure 10 *	1-6, 9, 14-22, 25, 30-34	C 12 N 15/00 A 01 H 1/00		
P, Y	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 261, no. 1, 5th January 1986, pages 108-121, The American Society of Biological Chemists, Inc.; J.L. SLIGHTOM et al.: "Nucleotide sequence analysis of TL-DNA of <i>Agrobacterium rhizogenes</i> agropine type plasmid" * Table V *	1-40			
P, Y	EP-A-0 145 338 (AGRIGENETICS) * Whole document *	7, 8, 10-13, 23, 24, 26-29, 37-40			
A	---	1-40	TECHNICAL FIELDS SEARCHED (Int. Cl.4)		
A	CHEMICAL ABSTRACTS, vol. 101, 1984, page 152, no. 123939z, Columbus, Ohio, US; D. TEPFER: "Transformation of several species of higher plants by <i>Agrobacterium rhizogenes</i> : sexual transmission of the transformed genotype and phenotype" & CELL (CAMBRIDGE, MASS.) 1984, 37(3), 959-67 * Whole abstract *	1-40	C 12 N A 01 H		
A	EP-A-0 126 546 (AGRIGENETICS) * Page 14, paragraph 2 - page 15; example 10 *	9-13, 25-29			
A	WO-A-8 402 913 (MONSANTO CO.) * Whole document *	1, 7, 17, 37			
	---	-/-			
The present search report has been drawn up for all claims					
Place of search	Date of completion of the search	Examiner			
THE HAGUE	03-05-1988	MADDOX A.D.			
CATEGORY OF CITED DOCUMENTS					
X : particularly relevant if taken alone	T : theory or principle underlying the invention				
Y : particularly relevant if combined with another document of the same category	E : earlier patent document, but published on, or after the filing date				
A : technological background	D : document cited in the application				
O : non-written disclosure	L : document cited for other reasons				
P : intermediate document	& : member of the same patent family, corresponding document				



European Patent
Office

EUROPEAN SEARCH REPORT

Page 0204590

Application Number

EP 86 40 0855

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl. 4)
			TECHNICAL FIELDS SEARCHED (Int. Cl. 4)
A	PLASMID, vol. 12, 1984, pages 91-102, Academic Press Inc.; L. JOUANIN: "Restriction map of an agropine-type Ri plasmid and its homologies with Ti plasmids" ---		
A	CHEMICAL ABSTRACTS, vol. 97, 1982, pages 203-204, no. 86183z, Columbus, Ohio, US; L. WILLMITZER et al.: "DNA from agrobacterium rhizogenes is transferred to and expressed in axenic hairy root plant tissues" & MGG, MOL. GEN. GENET. 1982, 186(1), 16-22 ---		
A	BIOLOGICAL ABSTRACTS/RRM, no. 30004915, Philadelphia, PA, US; R. PEERBOLTE et al.: "Hairy root T DNA T DNA transcripts and progeny" & ADVANCES IN AGRICULTURAL BIOTECHNOLOGY: GENETIC 90-247-3131-3. 1985. Vol. 0, no 0 p76-77 ---	1-6, 14, 17-22	
E	WO-A-8 700 865 (CALGENE INC.) * Claim 3 * -----	1, 9, 14	

The present search report has been drawn up for all claims

EPO FORM IS03 03.82 (1981)

Place of search	Date of completion of the search	Examiner
THE HAGUE	03-05-1988	MADDOX A.D.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons
& : member of the same patent family, corresponding document

THIS PAGE BLANK (USPTO)